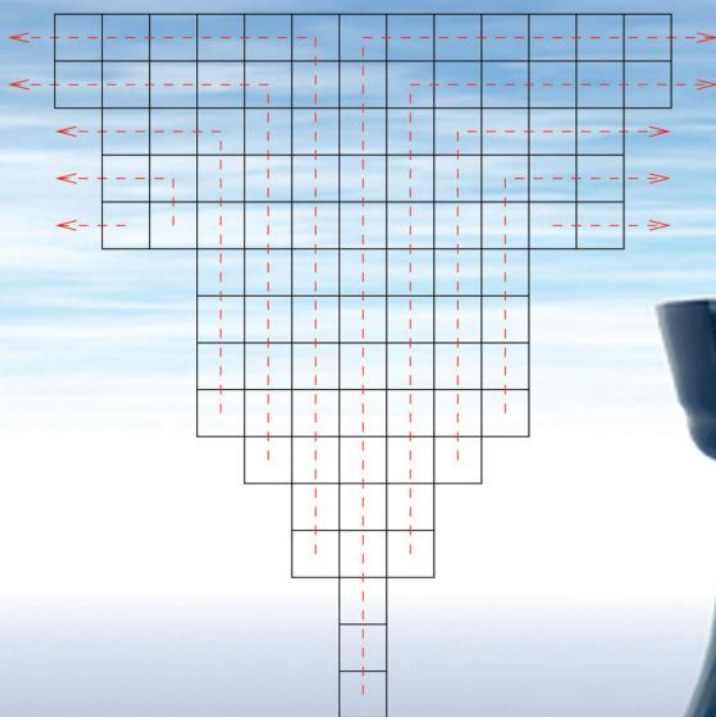


DISCRETE MATHEMATICS AND ITS APPLICATIONS

Series Editor KENNETH H. ROSEN

# BIJECTIVE COMBINATORICS



Nicholas A. Loehr



CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

# BIJECTIVE COMBINATORICS

# DISCRETE MATHEMATICS AND ITS APPLICATIONS

Series Editor  
Kenneth H. Rosen, Ph.D.

*R. B. J. T. Allenby and Alan Slomson, How to Count: An Introduction to Combinatorics, Third Edition*

*Juergen Bierbrauer, Introduction to Coding Theory*

*Donald Bindner and Martin Erickson, A Student's Guide to the Study, Practice, and Tools of Modern Mathematics*

*Francine Blanchet-Sadri, Algorithmic Combinatorics on Partial Words*

*Richard A. Brualdi and Dragoš Cvetković, A Combinatorial Approach to Matrix Theory and Its Applications*

*Kun-Mao Chao and Bang Ye Wu, Spanning Trees and Optimization Problems*

*Charalambos A. Charalambides, Enumerative Combinatorics*

*Gary Chartrand and Ping Zhang, Chromatic Graph Theory*

*Henri Cohen, Gerhard Frey, et al., Handbook of Elliptic and Hyperelliptic Curve Cryptography*

*Charles J. Colbourn and Jeffrey H. Dinitz, Handbook of Combinatorial Designs, Second Edition*

*Martin Erickson, Pearls of Discrete Mathematics*

*Martin Erickson and Anthony Vazzana, Introduction to Number Theory*

*Steven Furino, Ying Miao, and Jianxing Yin, Frames and Resolvable Designs: Uses, Constructions, and Existence*

*Mark S. Gockenbach, Finite-Dimensional Linear Algebra*

*Randy Goldberg and Lance Riek, A Practical Handbook of Speech Coders*

*Jacob E. Goodman and Joseph O'Rourke, Handbook of Discrete and Computational Geometry, Second Edition*

*Jonathan L. Gross, Combinatorial Methods with Computer Applications*

*Jonathan L. Gross and Jay Yellen, Graph Theory and Its Applications, Second Edition*

*Jonathan L. Gross and Jay Yellen, Handbook of Graph Theory*

*David S. Gunderson, Handbook of Mathematical Induction: Theory and Applications*

*Darrel R. Hankerson, Greg A. Harris, and Peter D. Johnson, Introduction to Information Theory and Data Compression, Second Edition*

*Darel W. Hardy, Fred Richman, and Carol L. Walker, Applied Algebra: Codes, Ciphers, and Discrete Algorithms, Second Edition*

## ***Titles (continued)***

- Daryl D. Harms, Miroslav Kraetzl, Charles J. Colbourn, and John S. Devitt*, Network Reliability: Experiments with a Symbolic Algebra Environment
- Silvia Heubach and Toufik Mansour*, Combinatorics of Compositions and Words
- Leslie Hogben*, Handbook of Linear Algebra
- Derek F. Holt with Bettina Eick and Eamonn A. O'Brien*, Handbook of Computational Group Theory
- David M. Jackson and Terry I. Visentin*, An Atlas of Smaller Maps in Orientable and Nonorientable Surfaces
- Richard E. Klima, Neil P. Sigmon, and Ernest L. Stitzinger*, Applications of Abstract Algebra with Maple™ and MATLAB®, Second Edition
- Patrick Knupp and Kambiz Salari*, Verification of Computer Codes in Computational Science and Engineering
- William Kocay and Donald L. Kreher*, Graphs, Algorithms, and Optimization
- Donald L. Kreher and Douglas R. Stinson*, Combinatorial Algorithms: Generation Enumeration and Search
- Hang T. Lau*, A Java Library of Graph Algorithms and Optimization
- C. C. Lindner and C. A. Rodger*, Design Theory, Second Edition
- Nicholas A. Loehr*, Bijective Combinatorics
- Elliott Mendelson*, Introduction to Mathematical Logic, Fifth Edition
- Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone*, Handbook of Applied Cryptography
- Richard A. Mollin*, Advanced Number Theory with Applications
- Richard A. Mollin*, Algebraic Number Theory, Second Edition
- Richard A. Mollin*, Codes: The Guide to Secrecy from Ancient to Modern Times
- Richard A. Mollin*, Fundamental Number Theory with Applications, Second Edition
- Richard A. Mollin*, An Introduction to Cryptography, Second Edition
- Richard A. Mollin*, Quadratics
- Richard A. Mollin*, RSA and Public-Key Cryptography
- Carlos J. Moreno and Samuel S. Wagstaff, Jr.*, Sums of Squares of Integers
- Dingyi Pei*, Authentication Codes and Combinatorial Designs
- Kenneth H. Rosen*, Handbook of Discrete and Combinatorial Mathematics
- Douglas R. Shier and K.T. Wallenius*, Applied Mathematical Modeling: A Multidisciplinary Approach
- Alexander Stanoyevitch*, Introduction to Cryptography with Mathematical Foundations and Computer Implementations
- Jörn Steuding*, Diophantine Analysis
- Douglas R. Stinson*, Cryptography: Theory and Practice, Third Edition
- Roberto Togneri and Christopher J. deSilva*, Fundamentals of Information Theory and Coding Design
- W. D. Wallis*, Introduction to Combinatorial Designs, Second Edition
- W. D. Wallis and J. C. George*, Introduction to Combinatorics
- Lawrence C. Washington*, Elliptic Curves: Number Theory and Cryptography, Second Edition

This page intentionally left blank

DISCRETE MATHEMATICS AND ITS APPLICATIONS

Series Editor KENNETH H. ROSEN

# BIJECTIVE COMBINATORICS

Nicholas A. Loehr

Virginia Tech  
Blacksburg, Virginia, USA



CRC Press

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

Chapman & Hall/CRC  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2011 by Taylor and Francis Group, LLC  
Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper  
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4398-4886-9 (Ebook-PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

**Visit the Taylor & Francis Web site at**  
**<http://www.taylorandfrancis.com>**

**and the CRC Press Web site at**  
**<http://www.crcpress.com>**

---

# Contents

---

<b>Preface</b>	<b>xiii</b>
<b>Epigraph</b>	<b>xv</b>
<b>Introduction</b>	<b>xvii</b>
<b>1 Basic Counting</b>	<b>1</b>
1.1 Review of Set Theory . . . . .	1
1.2 Sum Rule . . . . .	2
1.3 Product Rule . . . . .	3
1.4 Words, Permutations, and Subsets . . . . .	4
1.5 Functions . . . . .	6
1.6 Bijections, Cardinality, and Counting . . . . .	9
1.7 Subsets, Binary Words, and Compositions . . . . .	11
1.8 Subsets of a Fixed Size . . . . .	12
1.9 Anagrams . . . . .	14
1.10 Lattice Paths . . . . .	16
1.11 Multisets . . . . .	19
1.12 Probability . . . . .	21
1.13 Games of Chance . . . . .	24
1.14 Conditional Probability and Independence . . . . .	29
Summary . . . . .	32
Exercises . . . . .	33
<b>2 Combinatorial Identities and Recursions</b>	<b>41</b>
2.1 Generalized Distributive Law . . . . .	41
2.2 Multinomial and Binomial Theorems . . . . .	44
2.3 Combinatorial Proofs . . . . .	47
2.4 Recursions . . . . .	51
2.5 Recursions for Multisets and Anagrams . . . . .	55
2.6 Recursions for Lattice Paths . . . . .	57
2.7 Catalan Recursions . . . . .	61
2.8 Integer Partitions . . . . .	65
2.9 Set Partitions . . . . .	71
2.10 Surjections . . . . .	74
2.11 Stirling Numbers and Rook Theory . . . . .	75
2.12 Linear Algebra Review . . . . .	79
2.13 Stirling Numbers and Polynomials . . . . .	80
2.14 Combinatorial Proofs of Polynomial Identities . . . . .	84
Summary . . . . .	86
Exercises . . . . .	89



<b>3</b>	<b>Counting Problems in Graph Theory</b>	<b>97</b>
3.1	Graphs and Digraphs . . . . .	97
3.2	Walks and Matrices . . . . .	99
3.3	DAGs and Nilpotent Matrices . . . . .	102
3.4	Vertex Degrees . . . . .	105
3.5	Functional Digraphs . . . . .	107
3.6	Cycle Structure of Permutations . . . . .	109
3.7	Counting Rooted Trees . . . . .	111
3.8	Connectedness and Components . . . . .	113
3.9	Forests . . . . .	116
3.10	Trees . . . . .	117
3.11	Counting Trees . . . . .	119
3.12	Pruning Maps . . . . .	121
3.13	Ordered Trees and Terms . . . . .	123
3.14	Ordered Forests and Lists of Terms . . . . .	125
3.15	Graph Coloring . . . . .	127
3.16	Spanning Trees . . . . .	131
3.17	Matrix-Tree Theorem . . . . .	134
3.18	Eulerian Tours . . . . .	137
	Summary . . . . .	140
	Exercises . . . . .	143
<b>4</b>	<b>Inclusion-Exclusion and Related Techniques</b>	<b>153</b>
4.1	Involutions . . . . .	153
4.2	The Inclusion-Exclusion Formula . . . . .	156
4.3	More Proofs of Inclusion-Exclusion . . . . .	158
4.4	Applications of the Inclusion-Exclusion Formula . . . . .	160
4.5	Derangements . . . . .	163
4.6	Coefficients of Chromatic Polynomials . . . . .	165
4.7	Classical Möbius Inversion . . . . .	165
4.8	Partially Ordered Sets . . . . .	168
4.9	Möbius Inversion for Posets . . . . .	169
4.10	Product Posets . . . . .	172
	Summary . . . . .	173
	Exercises . . . . .	175
<b>5</b>	<b>Ranking and Unranking</b>	<b>181</b>
5.1	Ranking, Unranking, and Related Problems . . . . .	181
5.2	Bijective Sum Rule . . . . .	182
5.3	Bijective Product Rule . . . . .	183
5.4	Ranking Words . . . . .	187
5.5	Ranking Permutations . . . . .	189
5.6	Ranking Subsets . . . . .	190
5.7	Ranking Anagrams . . . . .	192
5.8	Ranking Integer Partitions . . . . .	193
5.9	Ranking Set Partitions . . . . .	194
5.10	Ranking Card Hands . . . . .	196
5.11	Ranking Dyck Paths . . . . .	199
5.12	Ranking Trees . . . . .	201
5.13	Successors and Predecessors . . . . .	201
5.14	Random Selection . . . . .	203

Summary . . . . .	205
Exercises . . . . .	206
<b>6 Counting Weighted Objects</b>	<b>213</b>
6.1 Weighted Sets . . . . .	213
6.2 Inversions . . . . .	216
6.3 Weight-Preserving Bijections . . . . .	217
6.4 Sum and Product Rules for Weighted Sets . . . . .	218
6.5 Inversions and Quantum Factorials . . . . .	220
6.6 Descents and Major Index . . . . .	221
6.7 Quantum Binomial Coefficients . . . . .	223
6.8 Quantum Multinomial Coefficients . . . . .	227
6.9 Foata's Map . . . . .	230
6.10 Quantum Catalan Numbers . . . . .	233
Summary . . . . .	235
Exercises . . . . .	237
<b>7 Formal Power Series</b>	<b>243</b>
7.1 The Ring of Formal Power Series . . . . .	244
7.2 Finite Products and Powers of Formal Series . . . . .	247
7.3 Formal Polynomials . . . . .	248
7.4 Order of Formal Power Series . . . . .	251
7.5 Formal Limits, Infinite Sums, and Infinite Products . . . . .	252
7.6 Multiplicative Inverses in $K[x]$ and $K[[x]]$ . . . . .	255
7.7 Formal Laurent Series . . . . .	257
7.8 Formal Derivatives . . . . .	259
7.9 Composition of Polynomials . . . . .	261
7.10 Composition of Formal Power Series . . . . .	262
7.11 Generalized Binomial Expansion . . . . .	265
7.12 Generalized Powers of Formal Series . . . . .	268
7.13 Partial Fraction Expansions . . . . .	270
7.14 Application to Recursions . . . . .	274
7.15 Formal Exponentiation and Formal Logarithms . . . . .	277
7.16 Multivariable Polynomials and Formal Series . . . . .	279
Summary . . . . .	281
Exercises . . . . .	284
<b>8 The Combinatorics of Formal Power Series</b>	<b>291</b>
8.1 Sum Rule for Infinite Weighted Sets . . . . .	291
8.2 Product Rule for Infinite Weighted Sets . . . . .	292
8.3 Generating Functions for Trees . . . . .	294
8.4 Compositional Inversion Formulas . . . . .	296
8.5 Generating Functions for Partitions . . . . .	298
8.6 Partition Bijections . . . . .	300
8.7 Euler's Pentagonal Number Theorem . . . . .	303
8.8 Stirling Numbers of the First Kind . . . . .	306
8.9 Stirling Numbers of the Second Kind . . . . .	307
8.10 The Exponential Formula . . . . .	309
Summary . . . . .	311
Exercises . . . . .	313

<b>9</b>	<b>Permutations and Group Actions</b>	<b>319</b>
9.1	Definition and Examples of Groups . . . . .	319
9.2	Basic Properties of Groups . . . . .	321
9.3	Notation for Permutations . . . . .	322
9.4	Inversions and Sign . . . . .	325
9.5	Determinants . . . . .	329
9.6	Multilinearity and Laplace Expansions . . . . .	331
9.7	Cauchy-Binet Formula . . . . .	334
9.8	Subgroups . . . . .	336
9.9	Automorphism Groups of Graphs . . . . .	338
9.10	Group Homomorphisms . . . . .	341
9.11	Group Actions . . . . .	344
9.12	Permutation Representations . . . . .	347
9.13	Stable Subsets and Orbits . . . . .	349
9.14	Cosets . . . . .	351
9.15	The Size of an Orbit . . . . .	354
9.16	Conjugacy Classes in $S_n$ . . . . .	356
9.17	Applications of the Orbit Size Formula . . . . .	357
9.18	The Number of Orbits . . . . .	359
9.19	Pólya's Formula . . . . .	362
	Summary . . . . .	364
	Exercises . . . . .	367
<b>10</b>	<b>Tableaux and Symmetric Polynomials</b>	<b>377</b>
10.1	Partition Diagrams and Skew Shapes . . . . .	377
10.2	Tableaux . . . . .	379
10.3	Schur Polynomials . . . . .	380
10.4	Symmetric Polynomials . . . . .	383
10.5	Homogeneous Symmetric Polynomials . . . . .	385
10.6	Symmetry of Schur Polynomials . . . . .	387
10.7	Orderings on Partitions . . . . .	389
10.8	Schur Bases . . . . .	392
10.9	Tableau Insertion . . . . .	394
10.10	Reverse Insertion . . . . .	397
10.11	Bumping Comparison Theorem . . . . .	399
10.12	Pieri Rules . . . . .	401
10.13	Schur Expansion of $h_\alpha$ . . . . .	403
10.14	Schur Expansion of $e_\alpha$ . . . . .	406
10.15	Algebraic Independence . . . . .	407
10.16	Power-Sum Symmetric Polynomials . . . . .	408
10.17	Relations between $e$ 's and $h$ 's . . . . .	410
10.18	Generating Functions for $e$ 's and $h$ 's . . . . .	411
10.19	Relations between $p$ 's, $e$ 's, and $h$ 's . . . . .	413
10.20	Power-Sum Expansion of $h_n$ and $e_n$ . . . . .	414
10.21	The Involution $\omega$ . . . . .	417
10.22	Permutations and Tableaux . . . . .	419
10.23	Words and Tableaux . . . . .	424
10.24	Matrices and Tableaux . . . . .	427
10.25	Cauchy Identities . . . . .	429
10.26	Dual Bases . . . . .	431
	Summary . . . . .	433

Exercises . . . . .	436
<b>11 Abaci and Antisymmetric Polynomials</b>	<b>445</b>
11.1 Abaci and Integer Partitions . . . . .	445
11.2 Jacobi Triple Product Identity . . . . .	447
11.3 Ribbons and $k$ -Cores . . . . .	448
11.4 $k$ -Quotients and Hooks . . . . .	454
11.5 Antisymmetric Polynomials . . . . .	457
11.6 Labeled Abaci . . . . .	460
11.7 Pieri Rule for $p_k$ . . . . .	462
11.8 Pieri Rule for $e_k$ . . . . .	465
11.9 Pieri Rule for $h_k$ . . . . .	467
11.10 Antisymmetric Polynomials and Schur Polynomials . . . . .	469
11.11 Rim-Hook Tableaux . . . . .	470
11.12 Abaci and Tableaux . . . . .	474
11.13 Skew Schur Polynomials . . . . .	477
11.14 Jacobi-Trudi Formulas . . . . .	478
11.15 Inverse Kostka Matrix . . . . .	482
11.16 Schur Expansion of Skew Schur Polynomials . . . . .	485
11.17 Products of Schur Polynomials . . . . .	489
Summary . . . . .	491
Exercises . . . . .	493
<b>12 Additional Topics</b>	<b>497</b>
12.1 Cyclic Shifting of Paths . . . . .	497
12.2 Chung-Feller Theorem . . . . .	499
12.3 Rook-Equivalence of Ferrers Boards . . . . .	503
12.4 Parking Functions . . . . .	505
12.5 Parking Functions and Trees . . . . .	507
12.6 Möbius Inversion and Field Theory . . . . .	511
12.7 Quantum Binomial Coefficients and Subspaces . . . . .	513
12.8 Tangent and Secant Numbers . . . . .	517
12.9 Tournaments and the Vandermonde Determinant . . . . .	520
12.10 Hook-Length Formula . . . . .	523
12.11 Knuth Equivalence . . . . .	527
12.12 Pfaffians and Perfect Matchings . . . . .	533
12.13 Domino Tilings of Rectangles . . . . .	539
Summary . . . . .	545
Exercises . . . . .	547
<b>Answers and Hints to Selected Exercises</b>	<b>555</b>
<b>Bibliography</b>	<b>569</b>
<b>Index</b>	<b>577</b>

This page intentionally left blank

---

# *Preface*

---

This book presents a general introduction to enumerative combinatorics that emphasizes bijective methods. The text contains a systematic development of the mathematical tools needed to solve enumeration problems: basic counting rules, recursions, inclusion-exclusion techniques, generating functions, bijective proofs, and linear-algebraic methods. These tools are used to analyze many combinatorial structures including words, permutations, subsets, functions, compositions, integer partitions, graphs, trees, lattice paths, multisets, rook placements, set partitions, Eulerian tours, derangements, posets, tilings, and abaci. Later chapters delve into some of the algebraic aspects of combinatorics, including detailed treatments of formal power series, symmetric groups, group actions, symmetric polynomials, determinants, and the combinatorial calculus of tableaux.

This text is suitable for enumerative combinatorics courses at the beginning graduate or advanced undergraduate levels. The book is somewhat more advanced than standard undergraduate texts on discrete mathematics, but is less mathematically demanding than the technical masterpieces in the subject (e.g., Stanley's two-volume treatise on enumeration [127] or Macdonald's monograph on symmetric functions [89]). There should be ample material in the book for a year-long course. A one-semester introduction to combinatorics might cover most of the first eight chapters, possibly excluding Chapter 3 if there is a separate course offered on graph theory. The more technical aspects of Chapter 7 (on formal power series) can be skipped or skimmed over if pressed for time. A course emphasizing abstract algebra and its applications to combinatorics could be based on Chapters 2, 7, 9, 10, and 11. Chapter 12 consists of independent sections on optional topics that complement material in the main text. In many chapters, some of the later sections can be omitted without loss of continuity.

In principle, the text requires no mathematical prerequisites except for a familiarity with basic logic, set theory, and proof techniques. Certain sections of the book (which can be skipped in more elementary courses) do assume the reader has had some exposure to ideas from linear algebra, such as linear independence and bases. The chapters dealing with abstract algebraic structures (groups, rings, fields, vector spaces, and formal power series) are self-contained, providing all relevant definitions as they are needed. Thus, students and scholars with no prior background in algebra or combinatorics can profitably use this book for reference or self-study.

Each chapter ends with a summary, a set of exercises, and bibliographic notes. The book contains nearly one thousand exercises, ranging in difficulty from routine verifications to unsolved problems. Solutions, hints, or partial answers to many of these exercises are given in an appendix. Although we provide references to the literature for some of the major theorems and harder problems, no attempt has been made to pinpoint the original source for every result appearing in the text and exercises.

I am grateful to several anonymous reviewers for valuable feedback and comments on an early version of the manuscript, and to Bob Stern and the other editors and staff at CRC Press for their facilitation of the publication process. The dedicated efforts of copyeditors and proofreaders removed many errors from this text, but I bear full responsibility for those that remain. Readers may communicate errors and other comments to the author by sending

e-mail to `nloehr@vt.edu`. Errata and other pertinent information will be maintained on the book's website: <http://www.math.vt.edu/people/nloehr/bijbook.html>

I thank E. Brown, H. Freeman, F. Loehr, L. Lopez, A. Mendes, and G. Warrington for their advice and support during the writing of this book. Finally, I wish to mention some very special people who died before this book could be completed: Julie, Elina, and 32 of my fellow students and faculty at Virginia Tech who were lost three years ago on this date.

*Nicholas A. Loehr*

April 16, 2010

## EPIGRAPH

“Meaningless! Meaningless!” says the Teacher.  
“Utterly meaningless! Everything is meaningless.”  
What is twisted cannot be straightened;  
what is lacking cannot be counted.

— Ecclesiastes 1:2,15.



This page intentionally left blank

---

# Introduction

---

*Enumerative combinatorics* is the mathematical theory of counting. How many ways can we deal a thirteen-card bridge hand that has three face cards and is void in clubs? How many functions map a ten element set onto a seven element set? How many rearrangements of  $1, 2, \dots, n$  have no decreasing subsequence of length three? How many ways can we divide an assembly of twenty people into five groups? How many invertible functions on  $\{1, 2, \dots, n\}$  are equal to their own inverse? How many ways can we seat ten men and five women at a circular table so no two women are adjacent? How many ways can we write a positive integer  $n$  as a sum of positive integers? The techniques of enumerative combinatorics allow us to find answers to questions like these.

This book develops the basic principles of enumeration, placing particular emphasis on the role of *bijective proofs*. To prove that a set  $S$  of objects has size  $n$  bijectively, one must construct an explicit one-to-one correspondence (bijection) from  $S$  onto the set  $\{1, 2, \dots, n\}$ . More generally, one can prove that two sets  $A$  and  $B$  have the same size by exhibiting a bijection between  $A$  and  $B$ . For example, fix  $n \geq 1$  and let  $A$  be the set of all strings  $w_1 w_2 \cdots w_{2n}$  consisting of  $n$  left parentheses and  $n$  right parentheses that are *balanced* (every left parenthesis can be matched to a right parenthesis later in the sequence). Let  $B$  be the set of all arrays

$$\begin{pmatrix} y_1 & y_2 & \cdots & y_n \\ z_1 & z_2 & \cdots & z_n \end{pmatrix}$$

such that every number in  $\{1, 2, \dots, 2n\}$  appears once in the array,  $y_1 < y_2 < \cdots < y_n$ ,  $z_1 < z_2 < \cdots < z_n$ , and  $y_i < z_i$  for every  $i$ . The sets  $A$  and  $B$  seem quite different at first glance. Yet, we can prove that  $A$  and  $B$  have the same cardinality by means of the following bijection. Given  $w = w_1 w_2 \cdots w_{2n} \in A$ , let  $y_1 < y_2 < \cdots < y_n$  be the positions of the left parentheses in  $w$  (taken in increasing order), and let  $z_1 < z_2 < \cdots < z_n$  be the positions of the right parentheses in  $w$  (in increasing order). For example, the string  $((())((()))())$  in  $A$  maps to the array

$$\begin{pmatrix} 1 & 2 & 4 & 7 & 8 & 9 & 13 \\ 3 & 5 & 6 & 10 & 11 & 12 & 14 \end{pmatrix}.$$

One may check that the requirement  $y_i < z_i$  for all  $i$  is equivalent to the fact that  $w$  is a *balanced* string of parentheses. The string  $w$  is uniquely determined by the array of  $y_i$ 's and  $z_i$ 's, and every such array arises from a suitable choice of  $w \in A$ . Thus we have defined the required one-to-one correspondence between  $A$  and  $B$ . We now know that the sets  $A$  and  $B$  have the same size, although we have not yet determined what that size is!

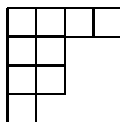
Bijective proofs, while elegant, can be very difficult to discover. For example, let  $C$  be the set of rearrangements of  $1, 2, \dots, n$  that have no decreasing subsequence of length three. There is a remarkable bijection between the set  $B$  (defined above) and the set  $C$ . But the reader may wish to defer a search for such a bijection until reading §12.11.

Luckily, the field of enumerative combinatorics contains a whole arsenal of techniques to help us solve complicated enumeration problems. Besides bijections, some of these techniques include recursions, generating functions, group actions, inclusion-exclusion formulas, linear algebra, probabilistic methods, symmetric polynomials, and more. We end this introduction by describing several challenging enumeration problems that can be solved using

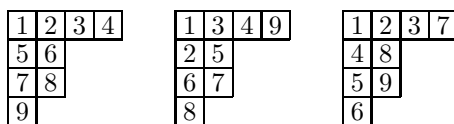
these more advanced methods. These problems, and the combinatorial technology needed to solve them, will be discussed at greater length later in the text.

## Standard Tableaux

Suppose we are given a diagram  $D$  consisting of a number of rows of boxes, left-justified, with each row no longer than the one above it. For example, consider the diagram:

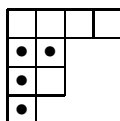


Let  $n$  be the total number of boxes in the diagram. A *standard tableau of shape  $D$*  is a filling of the boxes in  $D$  with the numbers  $1, 2, \dots, n$  (used once each) so that every row forms an increasing sequence (reading left to right), and every column forms an increasing sequence (reading top to bottom). For example, here are three standard tableaux of shape  $D$ , where  $D$  is the diagram pictured above:

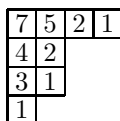


**Question:** *Given a diagram  $D$  of  $n$  cells, how many standard tableaux of shape  $D$  are there?*

There is a truly amazing answer to this counting problem, known as the *hook-length formula*. To state it, we need to define hooks and hook-lengths. The *hook* of a box  $b$  in a diagram  $D$  consists of all boxes to the right of  $b$  in its row, all boxes below  $b$  in its column, and box  $b$  itself. The *hook-length* of  $b$ , denoted  $h(b)$ , is the number of boxes in the hook of  $b$ . For example, if  $b$  is the first box in the second row of  $D$ , then the hook of  $b$  consists of the marked boxes in the following picture:



So  $h(b) = 4$ . In the picture below, we have labeled each box in  $D$  with its hook-length.



**Hook-Length Formula:** *The number of standard tableaux of shape  $D$  is  $n!$  divided by the product of the hook-lengths of all the boxes in  $D$ .*

For the diagram  $D$  in our example, the formula says there are exactly

$$\frac{9!}{7 \cdot 5 \cdot 2 \cdot 1 \cdot 4 \cdot 2 \cdot 3 \cdot 1 \cdot 1} = 216$$

standard tableaux of shape  $D$ . Observe that the set  $B$  of  $2 \times n$  arrays (discussed above) can

also be enumerated with the aid of the hook-length formula. In this case, the diagram  $D$  consists of two rows of length  $n$ . The hook-lengths for boxes in the top row are  $n+1, n, n-1, \dots, 2$ , while the hook-lengths in the bottom row are  $n, n-1, \dots, 1$ . Since there are  $2n$  boxes in all, the hook-length formula asserts that

$$|B| = \frac{(2n)!}{(n+1)n(n-1)\cdots 2 \cdot n(n-1)\cdots 1} = \frac{(2n)!}{(n+1)!n!}.$$

The fraction on the right side is an integer called the  $n$ th *Catalan number*. Since we previously displayed a bijection between  $B$  and  $A$  (the set of strings of balanced parentheses), we conclude that the size of  $A$  is also given by a Catalan number. As we will see, many different types of combinatorial structures are counted by the Catalan numbers.

How is the hook-length formula proved? Many proofs of this formula have been found since it was originally discovered in 1954. There are algebraic proofs, probabilistic proofs, combinatorial proofs, and (relatively recently) fully bijective proofs of this formula. Here we discuss a *flawed* probabilistic argument that gives a little intuition for how the mysterious hook-length formula arises. Suppose we choose a *random* filling  $F$  of the boxes of  $D$  with the integers  $1, 2, \dots, n$ . What is the probability that this filling will actually be a standard tableau? We remark that the filling is standard if and only if for every box  $b$  in  $D$ , the entry in  $b$  is the smallest number in the hook of  $b$ . Since any of the boxes in the hook is equally likely to contain the smallest value, we see that the probability of this event is  $1/h(b)$ . Multiplying these probabilities together would give  $1/\prod_{b \in D} h(b)$  as the probability that the random filling we chose is a standard tableau. Since the total number of possible fillings is  $n!$  (cf. Chapter 1), this leads us to the formula  $n!/\prod_{b \in D} h(b)$  for the number of standard tableaux of shape  $D$ .

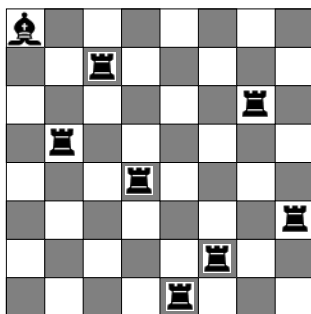
Unfortunately, the preceding argument contains a fatal error. The events “the entry in box  $b$  is the smallest in the hook of  $b$ ,” for various choices of  $b$ , are not necessarily *independent* (see §1.14). Thus we cannot find the probability that all these events occur by multiplying together the probabilities of each individual event. Nevertheless, remarkably, the final answer obtained by making this erroneous independence assumption turns out to be correct! This fact can be justified by a more subtle probabilistic argument due to Greene, Nijenhuis, and Wilf [62]. We describe this argument in §12.10.

## Rook Placements

A *rook* is a chess piece that can travel any number of squares along its current row or column in a single move. We say that the rook *attacks* all the squares in its row and column. How many ways can we place eight rooks on an ordinary  $8 \times 8$  chessboard so that no two rooks attack one another? The answer is  $8! = 40,320$ . More generally, we can show that there are  $n!$  ways to place  $n$  non-attacking rooks on an  $n \times n$  chessboard. To see this, first note that there must be exactly one rook in each of the  $n$  rows. The rook in the top row can occupy any of the  $n$  columns. The rook in the next row can occupy any of the  $n-1$  columns not attacked by the first rook; then there are  $n-2$  available columns for the next rook, and so on. By the product rule (discussed in Chapter 1), the total number of placements is therefore  $n \times (n-1) \times (n-2) \times \cdots \times 1 = n!$ .

Now consider an  $(n+1) \times (n+1)$  chessboard with a *bishop* occupying the upper-left corner square. (A bishop is a chess piece that attacks all squares that can be reached from its current square by moving in a straight line northeast, northwest, southeast, or southwest along a diagonal of the chessboard.) **Question:** *How many ways can we place  $n$  rooks on*

this chessboard so that no two pieces attack one another? An example of such a placement on a standard chessboard ( $n + 1 = 8$ , so  $n = 7$ ) is shown below:



It turns out that *the number of non-attacking placements is the closest integer to  $n!/e$* . Here,  $e$  is the famous constant  $e = \sum_{k=0}^{\infty} 1/k! \approx 2.718281828$  that appears throughout the subject of calculus. When  $n = 7$ , the number of placements is 1854 (note  $7!/e = 1854.112\ldots$ ).

This answer follows from the inclusion-exclusion formulas to be discussed in Chapter 4. We sketch the derivation now to indicate how the number  $e$  appears. First, there are  $n!$  ways to place the  $n$  rooks on the board so that no two rooks attack each other, and no rook occupies the top row or the leftmost column (lest a rook attack the bishop). However, we have counted many configurations in which one or more rooks occupy the diagonal attacked by the bishop. To correct for this, we will subtract a term that accounts for configurations of this kind. We can build such a configuration by placing a rook in row  $i$ , column  $i$ , for some  $i$  between 2 and  $n + 1$ , and then placing the remaining rooks in different rows and columns in  $(n - 1)!$  ways. So, presumably, we should subtract  $n \times (n - 1)! = n!$  from our original count of  $n!$ . But now our answer is zero! The trouble is that our subtracted term over-counts those configurations in which two or more rooks are attacked by the bishop. A naive count leads to the conclusion that there are  $\frac{n(n-1)}{2}(n-2)! = n!/2!$  such configurations, but this figure over-counts configurations with three or more rooks on the main diagonal. Thus we are led to a formula (called an inclusion-exclusion formula) in which we alternately add and subtract various terms to correct for all the over-counting. In the present situation, the final answer turns out to be

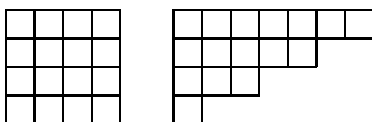
$$n! - n! + n!/2! - n!/3! + n!/4! - n!/5! + \cdots + (-1)^n n!/n! = n! \sum_{k=0}^n (-1)^k / k!.$$

Next, recall from calculus that  $e^x = \sum_{k=0}^{\infty} x^k / k!$  for all real  $x$ . In particular, taking  $x = -1$ , we have

$$e^{-1} = \frac{1}{e} = 1 - 1 + 1/2! - 1/3! + 1/4! - 1/5! + \cdots = \sum_{k=0}^{\infty} (-1)^k / k!.$$

We see that the combinatorial formula stated above consists of the first  $n + 1$  terms in the infinite series for  $n!/e$ . It can be shown (§4.5) that the “tail” of this series (namely  $\sum_{k=n+1}^{\infty} (-1)^k n!/k!$ ) is always less than 0.5 in absolute value. Thus, rounding  $n!/e$  to the nearest integer will produce the desired answer.

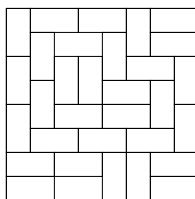
Another interesting combinatorial problem arises by comparing non-attacking rook placements on two boards of different shapes. For instance, consider the two generalized chessboards shown here:



One can check that for every  $k \geq 1$ , the number of ways to place  $k$  non-attacking rooks on the first board is the same as the number of ways to place  $k$  non-attacking rooks on the second board. We say that two boards are *rook-equivalent* whenever this property holds. It turns out that an  $n \times n$  board is always rook-equivalent to a board with successive row lengths  $2n - 1, 2n - 3, \dots, 5, 3, 1$ . More generally, there is a simple criterion for deciding whether two boards “of partition shape” are rook-equivalent. We will present this criterion in §12.3.

## Tilings

Now we turn to yet another problem involving chessboards. A *domino* is a rectangular object that can cover two horizontally or vertically adjacent squares on a chessboard. A *tiling* of a board is a covering of the board with dominos such that each square is covered by exactly one domino. For example, here is one possible tiling of a standard  $8 \times 8$  chessboard:



**Question:** Given a board of dimensions  $m \times n$ , how many ways can we tile it with dominos? This question may seem unfathomably difficult, so let us first consider the special case where  $m = 2$ . In this case, we are tiling a  $2 \times n$  region with dominos. Let  $f_n$  be the number of such tilings, for  $n = 0, 1, 2, \dots$ . One can see by drawing pictures that

$$f_0 = f_1 = 1, \quad f_2 = 2, \quad f_3 = 3, \quad f_4 = 5, \quad f_5 = 8, \quad f_6 = 13, \dots$$

The reader may recognize these numbers as being the start of the famous *Fibonacci sequence*. This sequence is defined recursively by letting  $F_0 = F_1 = 1$  and  $F_n = F_{n-1} + F_{n-2}$  for all  $n \geq 2$ . Now, a routine counting argument can be used to prove that the tiling numbers  $f_n$  satisfy the same recursive formula  $f_n = f_{n-1} + f_{n-2}$ . (To see this, note that a  $2 \times n$  tiling either ends with one vertical domino or two stacked horizontal dominos. Removing this part of the tiling either leaves a  $2 \times (n-1)$  tiling counted by  $f_{n-1}$  or a  $2 \times (n-2)$  tiling counted by  $f_{n-2}$ .) Since the sequences  $(f_n)$  and  $(F_n)$  satisfy the same recursion and initial conditions, they must agree for all  $n$ .

Now, what about the original tiling problem? Since the area of a tiled board must be even, there are no tilings unless at least one of the dimensions of the board is even. For boards satisfying this condition, Kasteleyn [75] and Fisher and Temperley [36] proved the following amazing result. *The number of domino tilings of an  $m \times n$  chessboard (with  $m$  even) is exactly equal to*

$$2^{mn/2} \prod_{j=1}^{m/2} \prod_{k=1}^n \sqrt{\cos^2 \left( \frac{j\pi}{m+1} \right) + \cos^2 \left( \frac{k\pi}{n+1} \right)}.$$

The formula is especially striking since the individual factors in the product are transcendental numbers, yet the product of all these factors is a positive integer! When  $m = n = 8$ , the formula reveals that the number of domino tilings of a standard chessboard is 12,988,816.

The proof of the formula involves *Pfaffians*, which are quantities analogous to determinants that arise in the study of skew-symmetric matrices. For details, see §12.12 and §12.13.

---

## Notes

Various proofs of the hook-length formula may be found in [42, 45, 62, 101, 108]. Treatments of rook-equivalence and other aspects of rook theory appear in [41, 55, 56, 74]. The domino tiling formula was proved by Kasteleyn [75] and discovered independently by Fisher and Temperley [36].

# Basic Counting

This chapter develops the basic counting techniques that form the foundation of enumerative combinatorics. We apply these techniques to study fundamental combinatorial structures such as words, permutations, subsets, functions, and lattice paths. The end of the chapter gives some applications of combinatorics to probability theory.

## 1.1 Review of Set Theory

We assume the reader is familiar with elementary aspects of logic and set theory, including proofs by induction. This material may be found in texts such as [34, 126]. Table 1.1 reviews the notation we will use from set theory. The word *iff* is defined to mean “if and only if.”

**TABLE 1.1**

Review of notation from set theory.

Concept	Symbol	Meaning
membership	$x \in S$	$x$ is an element of the set $S$ .
set-building	$\{x : P(x)\}$	$y \in \{x : P(x)\}$ iff $P(y)$ is true.
subset	$A \subseteq B$	For all $x$ , $x \in A$ implies $x \in B$ .
set equality	$A = B$	For all $x$ , $x \in A$ iff $x \in B$ .
empty set	$\emptyset$	For all $x$ , $x \notin \emptyset$ .
cardinality	$ A  = n$	The set $A$ has exactly $n$ members.
union	$A \cup B$	$x \in A \cup B$ iff $x \in A$ or $x \in B$ .
intersection	$A \cap B$	$x \in A \cap B$ iff $x \in A$ and $x \in B$ .
set difference	$A \sim B$	$x \in A \sim B$ iff $x \in A$ and $x \notin B$ .
ordered pair	$(a, b)$	$(a, b) = (c, d)$ iff $a = c$ and $b = d$ .
Cartesian product	$A \times B$	$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$ .
finite union	$A_1 \cup \cdots \cup A_n$	$x \in A_1 \cup \cdots \cup A_n$ iff $x \in A_i$ for at least one $i \leq n$ .
finite intersection	$A_1 \cap \cdots \cap A_n$	$x \in A_1 \cap \cdots \cap A_n$ iff $x \in A_i$ for all $i \leq n$ .
ordered $n$ -tuple	$(a_1, \dots, a_n)$	$(a_1, \dots, a_n) = (b_1, \dots, b_n)$ iff $a_i = b_i$ for $1 \leq i \leq n$ .
finite product	$A_1 \times \cdots \times A_n$	$A_1 \times \cdots \times A_n = \{(a_1, \dots, a_n) : a_i \in A_i \text{ for } i \leq n\}$ .

We use the notation  $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ ,  $\mathbb{N}^+ = \{1, 2, 3, \dots\}$ ,  $\mathbb{Z}$  for the set of all integers,  $\mathbb{Q}$  for the set of rational numbers,  $\mathbb{R}$  for the set of real numbers, and  $\mathbb{C}$  for the set of complex numbers. Informally, the notation  $|A| = n$  means that  $A$  is a set consisting of  $n$  elements. We will give a more formal discussion of cardinality later (§1.6).

Two sets  $A$  and  $B$  are *disjoint* iff  $A \cap B = \emptyset$ . More generally, the sets  $A_1, \dots, A_n$  are called *pairwise disjoint* iff  $A_i \cap A_j = \emptyset$  for all  $i \neq j$ . This means that no two sets in the given list overlap one another.



## 1.2 Sum Rule

The starting point for enumerative combinatorics is the following basic fact.

**1.1. Counting Principle.** If  $A$  and  $B$  are finite disjoint sets, then  $|A \cup B| = |A| + |B|$ .

The requirement that  $A$  and  $B$  be *disjoint* is certainly necessary. For example, if  $A = \{1, 2, 3\}$  and  $B = \{3, 5\}$ , then  $|A \cup B| = 4$ , while  $|A| + |B| = 3 + 2 = 5$ . We will give a formal proof of 1.1 later (see 1.32). For now, let us deduce some consequences of this counting principle.

**1.2. Sum Rule.** If  $A_1, \dots, A_m$  are pairwise disjoint finite sets, then

$$|A_1 \cup \dots \cup A_m| = |A_1| + \dots + |A_m|.$$

*Proof.* We use induction on  $m$ . The case  $m = 1$  is immediate, while the case  $m = 2$  is true by 1.1. For  $m > 2$ , assume the result is known for  $m - 1$  sets. In 1.1, let  $A = A_1 \cup \dots \cup A_{m-1}$  and  $B = A_m$ . By induction hypothesis,

$$|A| = |A_1| + \dots + |A_{m-1}|.$$

Since  $A_m$  does not intersect any  $A_j$  with  $j < m$ , we see that  $A$  and  $B$  are disjoint. So 1.1 gives

$$|A_1 \cup \dots \cup A_m| = |A \cup B| = |A| + |B| = |A_1| + \dots + |A_{m-1}| + |A_m|. \quad \square$$

**1.3. Difference Rule.** If  $S$  and  $T$  are finite sets such that  $T \subseteq S$ , then  $|S \sim T| = |S| - |T|$ .

*Proof.* The set  $S$  is the union of the disjoint sets  $T$  and  $S \sim T$ . Therefore, 1.1 gives  $|S| = |T| + |S \sim T|$ . Subtracting the finite quantity  $|T|$  from both sides gives the result.  $\square$

We can generalize 1.1 to the case where the two sets in question are not disjoint, as follows.

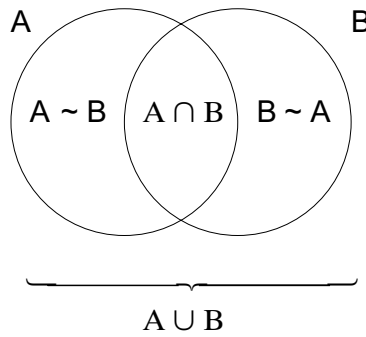
**1.4. Binary Union Rule.** If  $A$  and  $B$  are arbitrary finite sets, then  $|A \cup B| = |A| + |B| - |A \cap B|$ .

*Proof.* Note that  $A$  is the disjoint union of  $A \sim B$  and  $A \cap B$ ;  $B$  is the disjoint union of  $B \sim A$  and  $A \cap B$ ; and  $A \cup B$  is the disjoint union of  $A \sim B$ ,  $B \sim A$ , and  $A \cap B$ . See Figure 1.1. Applying the sum rule repeatedly, we see that

$$|A| = |A \sim B| + |A \cap B|; \quad |B| = |B \sim A| + |A \cap B|; \quad |A \cup B| = |A \sim B| + |B \sim A| + |A \cap B|.$$

Using the first two equations to eliminate  $|A \sim B|$  and  $|B \sim A|$  in the third equation, we obtain the desired result.  $\square$

The sum rule can also be extended to a formula for  $|A_1 \cup \dots \cup A_n|$ , where  $A_1, \dots, A_n$  are arbitrary (not necessarily pairwise disjoint) finite sets. This formula is called the *inclusion-exclusion formula*; we will study it later (Chapter 4).

**FIGURE 1.1**

Proof of the binary union rule.

### 1.3 Product Rule

We can use the sum rule to compute the size of the Cartesian product of finite sets.

**1.5. Product Rule for Sets.** Suppose  $S_1, \dots, S_k$  are finite sets with  $|S_i| = n_i$  for  $1 \leq i \leq k$ . Then

$$|S_1 \times S_2 \times \cdots \times S_k| = n_1 n_2 \cdots n_k.$$

*Proof.* We proceed by induction on  $k$ . There is nothing to prove when  $k = 1$ . Consider the case  $k = 2$ . Let  $S_2 = \{x_1, x_2, \dots, x_{n_2}\}$ . The set  $S_1 \times S_2$  is the disjoint union of the  $n_2$  sets  $S_1 \times \{x_1\}, S_1 \times \{x_2\}, \dots, S_1 \times \{x_{n_2}\}$ . Each of these sets has cardinality  $|S_1| = n_1$ . So, by the sum rule,

$$|S_1 \times S_2| = \sum_{i=1}^{n_2} |S_1 \times \{x_i\}| = \sum_{i=1}^{n_2} n_1 = n_1 n_2.$$

Next, let  $k > 2$  and assume that the result is already known for products of  $k - 1$  sets. We can regard  $S_1 \times S_2 \times \cdots \times S_k$  as the Cartesian product  $A \times B$ , where  $A = S_1 \times \cdots \times S_{k-1}$  and  $B = S_k$ . By induction,  $|A| = n_1 n_2 \cdots n_{k-1}$ . By the  $k = 2$  case,

$$|S_1 \times S_2 \times \cdots \times S_k| = |A \times B| = |A| \cdot |B| = (n_1 n_2 \cdots n_{k-1}) n_k.$$

This completes the induction step. □

**1.6. Example: License Plates.** A California license plate consists of a digit, followed by three uppercase letters, followed by three more digits. Formally, we can view a license plate as an element of the set  $S = D \times L \times L \times L \times D \times D \times D$ , where  $D = \{0, 1, 2, \dots, 9\}$  and  $L = \{A, B, C, \dots, Z\}$ . Thus,

$$|S| = 10 \times 26 \times 26 \times 26 \times 10 \times 10 \times 10 = 175,760,000.$$

**1.7. Example: Phone Numbers.** A phone number is a ten-digit sequence such that the first digit is not zero or one, while the second digit must be zero or one. Formally, we can view a phone number as an element of the set  $S = \{2, 3, \dots, 9\} \times \{0, 1\} \times D^8$ , where the notation  $D^8$  denotes the Cartesian product of 8 copies of the set  $D = \{0, 1, \dots, 9\}$ . The number of phone numbers is

$$|S| = 8 \times 2 \times 10^8 = 1.6 \text{ billion.}$$

(To allow for more phone numbers, the restriction on the second digit of the area code was removed years ago.)

We will often be interested in finding the cardinality of a finite set  $S$  whose members are “structured objects.” Frequently, we will be able to build up each object in  $S$  by making a sequence of choices. The next counting principle tells us how to compute  $|S|$  in this situation.

**1.8. Product Rule.** Suppose each object  $x$  in a set  $S$  can be uniquely constructed by making a sequence of  $k$  choices. Suppose the first choice can be made in  $n_1$  ways; the second choice can be made in  $n_2$  ways (regardless of what the first choice was); and so on. In general, we suppose that the  $i$ th choice can be made in  $n_i$  ways, regardless of what happened in the first  $i - 1$  choices, for all  $i \leq k$ . Then

$$|S| = n_1 n_2 \cdots n_k.$$

The product rule is a consequence of 1.5, as we will explain in 1.34.

**1.9. Example: Fraternity and Sorority Names.** The name of a fraternity or sorority consists of any sequence of two or three uppercase Greek letters. (The Greek alphabet has 24 letters.) How many possible names are there? The set  $S$  of all such names is the disjoint union of  $S_2$  and  $S_3$ , where  $S_k$  is the set of names of length  $k$ . Using the sum rule,

$$|S| = |S_2| + |S_3|.$$

We can calculate  $|S_2|$  using the product rule. We build a typical word in  $S_2$  by choosing the first letter (24 ways), then choosing the second letter (24 ways). By the product rule,  $|S_2| = 24^2$ . Similarly,  $|S_3| = 24^3$ , so  $|S| = 24^2 + 24^3 = 14,400$ . Note that we cannot directly use the product rule to calculate  $|S|$ , since the *number* of choices in a given application of the product rule must be fixed.

**1.10. Example.** How many three-digit odd numbers contain the digit 2 but not the digit 5? Let  $X$  be the set of all such numbers. We can write  $X$  as the disjoint union of three sets  $A$ ,  $B$ , and  $C$ , where  $A$  consists of numbers in  $X$  with first and second digit 2,  $B$  consists of numbers in  $X$  with first digit 2 and second digit not 2, and  $C$  consists of numbers in  $X$  with second digit 2 and first digit not 2. To build a number in  $C$ , we choose the digits from left to right. There are seven choices for the first digit (we must avoid 0, 2, and 5), one choice for the second digit (it must be 2), and four choices for the third digit (which is odd and unequal to 5). By the product rule,  $|C| = 7 \cdot 1 \cdot 4 = 28$ . Similar reasoning shows that  $|A| = 4$  and  $|B| = 1 \cdot 8 \cdot 4 = 32$ . Therefore,  $|X| = |A| + |B| + |C| = 64$ .

## 1.4 Words, Permutations, and Subsets

**1.11. Definition: Words.** Let  $A$  be a finite set. A *word* over the alphabet  $A$  is a sequence  $w = w_1 w_2 \cdots w_k$ , where each  $w_i \in A$  and  $k \geq 0$ . The *length* of  $w = w_1 w_2 \cdots w_k$  is  $k$ . Two words  $w = w_1 w_2 \cdots w_k$  and  $z = z_1 z_2 \cdots z_m$  are *equal* iff  $k = m$  and  $w_i = z_i$  for  $1 \leq i \leq k$ .

**1.12. Example.** Let  $A = \{a, b, c, \dots, z\}$  be the set of 26 lowercase letters in the English alphabet. Then *stop*, *opts*, and *stoops* are distinct words (of lengths 4, 4, and 6, respectively). If  $A = \{0, 1\}$ , the 8 words of length 3 over  $A$  are

$$000, \quad 001, \quad 010, \quad 011, \quad 100, \quad 101, \quad 110, \quad 111.$$

There is exactly one word of length zero, called the *empty word*. It is sometimes denoted by the special symbols  $\cdot$  or  $\epsilon$ .

**1.13. Theorem: Enumeration of Words.** If  $A$  is an  $n$ -letter alphabet and  $k \geq 0$ , then there are  $n^k$  words of length  $k$  over  $A$ .

*Proof.* We can uniquely construct a typical word  $w = w_1 w_2 \cdots w_k$  by a sequence of choices. First, choose  $w_1 \in A$  to be any of the  $n$  letters in  $A$ . Second, choose  $w_2 \in A$  in any of  $n$  ways. Continue similarly, choosing  $w_i \in A$  in any of  $n$  ways for  $1 \leq i \leq k$ . By the product rule, the number of words is  $n \times n \times \cdots \times n$  ( $k$  factors), which is  $n^k$ . Note that the empty word is the unique word of length 0 over  $A$ , so our formula holds for  $k = 0$  also.  $\square$

**1.14. Definition: Permutations.** Let  $A$  be an  $n$ -element set. A *permutation* of  $A$  is a word  $w = w_1 w_2 \cdots w_n$  in which each letter of  $A$  appears exactly once. For example, the 6 permutations of  $A = \{x, y, z\}$  are

$$xyz, \quad xzy, \quad yxz, \quad yzx, \quad zxy, \quad zyx.$$

**1.15. Definition: Factorials.** For each integer  $n \geq 1$ ,  $n$ -factorial is

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1,$$

which is the product of the first  $n$  positive integers. We also define  $0! = 1$ .

**1.16. Theorem: Enumeration of Permutations.** There are  $n!$  permutations of an  $n$ -letter alphabet  $A$ .

*Proof.* Build a typical permutation  $w = w_1 w_2 \cdots w_n$  of  $A$  by making  $n$  choices. First, choose  $w_1$  to be any of the  $n$  letters of  $A$ . Second, choose  $w_2$  to be any of the  $n-1$  letters of  $A$  different from  $w_1$ . Third, choose  $w_3$  to be any of the  $n-2$  letters of  $A$  different from  $w_1$  and  $w_2$ . Proceed similarly; at the  $n$ th stage, choose  $w_n$  to be the unique letter of  $A$  that is different from  $w_1, w_2, \dots, w_{n-1}$ . By the product rule, the number of permutations is  $n \times (n-1) \times \cdots \times 1 = n!$ . The result also holds when  $n = 0$ .  $\square$

**1.17. Definition:  $k$ -Permutations.** Let  $A$  be an  $n$ -element set. A  $k$ -permutation of  $A$  is a word  $w = w_1 w_2 \cdots w_k$  consisting of  $k$  distinct letters in  $A$ . For example, the twelve 2-permutations of  $A = \{a, b, c, d\}$  are

$$ab, \quad ac, \quad ad, \quad ba, \quad bc, \quad bd, \quad ca, \quad cb, \quad cd, \quad da, \quad db, \quad dc.$$

An  $n$ -permutation of  $A$  is the same as a permutation of  $A$ .

**1.18. Theorem: Enumeration of  $k$ -Permutations.** Suppose  $A$  is an  $n$ -letter alphabet. For  $0 \leq k \leq n$ , the number of  $k$ -permutations of  $A$  is

$$n(n-1)(n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}.$$

For  $k > n$ , there are no  $k$ -permutations of  $A$ .

*Proof.* Build a typical  $k$ -permutation  $w = w_1 w_2 \cdots w_k$  of  $A$  by making  $k$  choices. First, choose  $w_1$  to be any of the  $n$  letters of  $A$ . Second, choose  $w_2$  to be any of the  $n-1$  letters of  $A$  different from  $w_1$ . Continue similarly. When we choose  $w_i$  (where  $1 \leq i \leq k$ ), we have already used the  $i-1$  distinct letters  $w_1, w_2, \dots, w_{i-1}$ . Since  $A$  has  $n$  letters, there are  $n - (i-1) = n - i + 1$  choices available at stage  $i$ . In particular, for the  $k$ th and final choice, there are  $n - k + 1$  ways to choose  $w_k$ . By the product rule, the number of  $k$ -permutations is  $\prod_{i=1}^k (n - (i-1)) = n(n-1) \cdots (n-k+1)$ . Multiplying this expression by  $(n-k)!/(n-k)!$ , we obtain the product of the integers 1 through  $n$  in the numerator, which is  $n!$ . Thus the answer is also given by the formula  $n!/(n-k)!$ .  $\square$

**1.19. Definition: Power Set.** For any set  $S$ , the *power set*  $\mathcal{P}(S)$  is the set of all subsets of  $S$ . Thus,  $T \in \mathcal{P}(S)$  iff  $T \subseteq S$ . For example, if  $S = \{2, 5, 7\}$ , then  $\mathcal{P}(S)$  is the eight-element set

$$\{\emptyset, \{2\}, \{5\}, \{7\}, \{2, 5\}, \{2, 7\}, \{5, 7\}, \{2, 5, 7\}\}.$$

**1.20. Theorem: Cardinality of Power Sets.** An  $n$ -element set has  $2^n$  subsets. In other words, if  $|S| = n$ , then  $|\mathcal{P}(S)| = 2^n$ .

*Proof.* Suppose  $S = \{x_1, \dots, x_n\}$  is an  $n$ -element set. We can build a typical subset  $T$  of  $S$  by making a sequence of  $n$  choices. First, decide whether  $x_1 \in T$  or  $x_1 \notin T$ . This binary decision can be made in two ways. Second, decide whether  $x_2 \in T$  or  $x_2 \notin T$ ; again there are two possibilities. Continue similarly; decide in the  $i$ th choice whether  $x_i \in T$  or  $x_i \notin T$  (two possibilities). This sequence of choices uniquely determines which  $x_j$ 's belong to  $T$ . Since  $T$  is a subset of  $S$ , this information uniquely determines the set  $T$ . By the product rule, the number of subsets is  $2 \times 2 \times \dots \times 2$  ( $n$  factors), which is  $2^n$ .  $\square$

## 1.5 Functions

This section reviews the definitions of functions, injections, surjections, and bijections, which should already be familiar to the reader. We also enumerate the number of functions, injections, and bijections between two given finite sets. The enumeration of surjections is more subtle, and will be discussed later (§2.10).

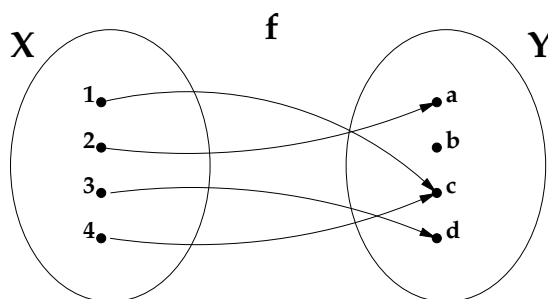
**1.21. Definition: Functions.** Formally, a *function*  $f$  from  $X$  to  $Y$  is an ordered triple  $(X, Y, G)$ , where  $G$  is a subset of  $X \times Y$  such that for each  $x \in X$  there is exactly one  $y \in Y$  with  $(x, y) \in G$ .  $X$  is the *domain* of  $f$ ,  $Y$  is the *codomain* of  $f$ , and  $G$  is the *graph* of  $f$ . We write  $y = f(x)$  iff  $(x, y) \in G$ , and we write  $f : X \rightarrow Y$  to signify that  $f$  is a function from  $X$  to  $Y$ . Let  ${}^XY$  denote the set of all functions from  $X$  to  $Y$ .

Informally, we think of a function  $f$  as consisting of a rule that maps each  $x \in X$  to a unique value  $f(x) \in Y$ . When  $X$  and  $Y$  are finite sets, it is convenient to visualize  $f$  by an *arrow diagram*. We obtain this diagram by drawing a dot for each element of  $X$  and  $Y$ , and drawing an arrow from  $x$  to  $y$  whenever  $y = f(x)$ . The definition of a function requires that each  $x \in X$  have *exactly one* arrow emanating from it, and the arrow must point to an element of  $Y$ . On the other hand, an element  $y \in Y$  may have zero, one, or more than one arrow hitting it. Figures 1.2, 1.3, 1.4, and 1.5 depict the arrow diagrams for some functions.

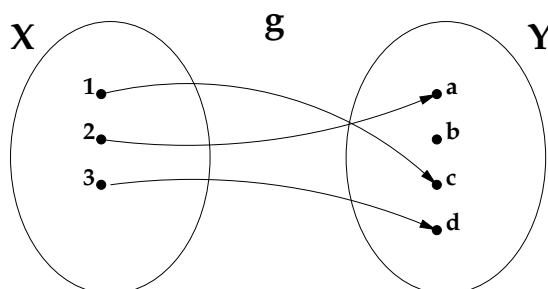
**1.22. Theorem: Enumeration of Functions.** Suppose  $X = \{x_1, \dots, x_n\}$  is an  $n$ -element set and  $Y = \{y_1, \dots, y_m\}$  is an  $m$ -element set. There are  $m^n$  functions from  $X$  to  $Y$ . In other words,  $|{}^XY| = |Y|^{|X|}$ .

*Proof.* To build a typical function  $f \in {}^XY$ , we make a sequence of  $n$  choices that uniquely determine the graph  $G$  of  $f$ . First, we choose  $f(x_1)$  to be any of the  $m$  elements of  $Y$ . Second, we choose  $f(x_2)$  to be any of the  $m$  elements of  $Y$ . Similarly, for each  $i \leq n$ , we choose  $f(x_i)$  to be any of the  $m$  elements of  $Y$ . By the product rule, the number of functions we can build is  $m \times m \times \dots \times m$  ( $n$  factors), which is  $m^n$ .  $\square$

**1.23. Definition: Injections.** A function  $g : X \rightarrow Y$  is an *injection* iff for all  $x, x' \in X$ ,  $x \neq x'$  implies  $g(x) \neq g(x')$ . Injective functions are also called *one-to-one* functions.

**FIGURE 1.2**

A function  $f : \{1, 2, 3, 4\} \rightarrow \{a, b, c, d\}$ .

**FIGURE 1.3**

An injective function  $g : \{1, 2, 3\} \rightarrow \{a, b, c, d\}$ .

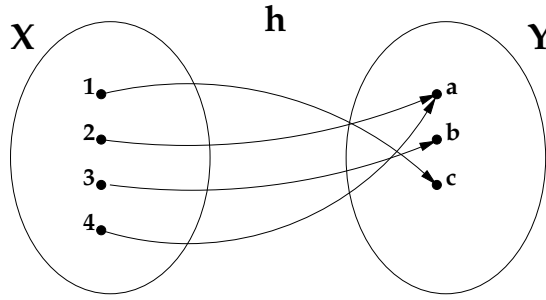
In the arrow diagram for an injective function, every  $y \in Y$  has *at most one* arrow entering it. For example, the function  $f$  in Figure 1.2 is not injective, while the function  $g$  in Figure 1.3 is injective.

**1.24. Theorem: Enumeration of Injections.** Suppose  $X = \{x_1, \dots, x_n\}$  is an  $n$ -element set and  $Y = \{y_1, \dots, y_m\}$  is an  $m$ -element set. If  $n \leq m$ , the number of injections from  $X$  into  $Y$  is  $m(m-1)(m-2) \cdots (m-n+1) = m!/(m-n)!$ . If  $n > m$ , there are no injections from  $X$  to  $Y$ .

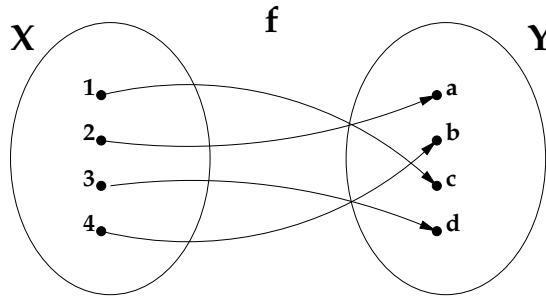
*Proof.* Assume first that  $n \leq m$ . As above, we construct a typical injection  $g : X \rightarrow Y$  by choosing the  $n$  function values  $g(x_i)$ , for  $1 \leq i \leq n$ . For each  $i \leq n$ , we choose  $g(x_i)$  to be an element of  $Y$  distinct from the elements  $g(x_1), \dots, g(x_{i-1})$  already chosen. Since the latter elements are pairwise distinct, we see that there are  $m - (i-1) = m - i + 1$  alternatives for  $g(x_i)$ , no matter what happened in the first  $i-1$  choices. By the product rule, the number of injections is  $m(m-1) \cdots (m-n+1) = m!/(m-n)!$ .

On the other hand, suppose  $n > m$ . Try to build an injection  $g$  by choosing the values  $g(x_1), g(x_2), \dots$  as before. When we try to choose  $g(x_{m+1})$ , there are no elements of  $Y$  distinct from the previously chosen elements  $g(x_1), \dots, g(x_m)$ . Since it is impossible to complete the construction of  $g$ , there are no injections from  $X$  to  $Y$  in this situation.  $\square$

**1.25. Definition: Surjections.** A function  $h : X \rightarrow Y$  is a *surjection* iff for every  $y \in Y$  there exists  $x \in X$  with  $y = f(x)$ . Surjective functions are also said to be *onto* or to map *onto* the codomain  $Y$ .

**FIGURE 1.4**

A surjective function  $h : \{1, 2, 3, 4\} \rightarrow \{a, b, c\}$ .

**FIGURE 1.5**

A bijective function  $f : \{1, 2, 3, 4\} \rightarrow \{a, b, c, d\}$ .

In the arrow diagram for a surjective function, every  $y \in Y$  has *at least one* arrow entering it. For example, the functions  $f$  and  $g$  in Figures 1.2 and 1.3 are not surjective, while the function  $h$  in Figure 1.4 is surjective. Note that  $h$  is not injective. Counting surjections is harder than counting other classes of functions, so we defer discussion of this problem to a later section (§2.10).

**1.26. Definition: Bijections.** A function  $f : X \rightarrow Y$  is a *bijection* iff  $f$  is both injective and surjective iff for every  $y \in Y$  there exists a unique  $x \in X$  with  $y = f(x)$ .

In the arrow diagram for a bijective function, every  $y \in Y$  has *exactly one* arrow entering it. For example, the functions in Figures 1.2 through 1.4 are not bijective, while the function  $f$  in Figure 1.5 is bijective.

**1.27. Theorem: Injectivity vs. Surjectivity.** Suppose  $f : X \rightarrow Y$  is a function. If  $X$  and  $Y$  are finite sets with the same number of elements, then  $f$  is injective iff  $f$  is surjective.

*Proof.* Suppose  $X$  and  $Y$  both have  $n$  elements, and write  $X = \{x_1, \dots, x_n\}$ . Assume that  $f : X \rightarrow Y$  is injective. Then the set  $T = \{f(x_1), \dots, f(x_n)\}$  is a subset of  $Y$  consisting of  $n$  *distinct* elements. Since  $Y$  has  $n$  elements, this subset must be all of  $Y$ . This means that every  $y \in Y$  has the form  $f(x_i)$  for some  $x_i \in X$ , so that  $f$  is surjective.

Conversely, assume that  $f : X \rightarrow Y$  is not injective. Then there exist  $i \neq j$  with  $f(x_i) = f(x_j)$ . It follows that the set  $T = \{f(x_1), \dots, f(x_n)\}$  contains fewer than  $n$  elements, since the displayed list of members of  $T$  contains at least one duplicate. Thus  $T$  is a proper subset of  $Y$ . Letting  $y$  be any element of  $Y \setminus T$ , we see that  $y$  does not have the form  $f(x)$  for any  $x \in X$ . Therefore  $f$  is not surjective.  $\square$

The previous result does not extend to infinite sets, as shown by the following examples. The function  $f : \mathbb{N} \rightarrow \mathbb{N}$  defined by  $f(n) = n + 1$  is injective but not surjective. The function  $g : \mathbb{N} \rightarrow \mathbb{N}$  defined by  $g(2k) = g(2k + 1) = k$  for all  $k \geq 0$  is surjective but not injective. The function  $\exp : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $\exp(x) = e^x$  is injective but not surjective. The function  $h : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $h(x) = x(x - 1)(x + 1)$  is surjective but not injective.

**1.28. Theorem: Enumeration of Bijections.** Suppose  $X$  and  $Y$  are two  $n$ -element sets. Then there are  $n!$  bijections from  $X$  to  $Y$ .

*Proof.* By 1.27, a function  $f : X \rightarrow Y$  is injective iff  $f$  is surjective. Therefore, under the assumption that  $|X| = |Y| = n$ ,  $f$  is injective iff  $f$  is bijective. We have already seen that the number of injections from  $X$  to  $Y$  is  $n!/(n - n)! = n!$ . The result follows.  $\square$

If  $X$  is an  $n$ -element set and  $Y$  is an  $m$ -element set and  $m \neq n$ , there are no bijections from  $X$  to  $Y$  (cf. the next section).

**1.29. Remark.** The reader may note the similarity between the formulas obtained here for functions and the formulas obtained earlier for words and permutations. This is not a coincidence. Indeed, we can formally define a word  $w_1 w_2 \cdots w_k$  over an alphabet  $A$  as the function  $w : \{1, 2, \dots, k\} \rightarrow A$  defined by  $w(i) = w_i$ . The number of such words (functions) is  $|A|^k$ . The word  $w_1 w_2 \cdots w_k$  is a  $k$ -permutation of  $A$  iff the  $w_i$ 's are all distinct iff  $w$  is an *injective* function. The word  $w_1 w_2 \cdots w_k$  is a permutation of  $A$  iff  $w$  is a *bijective* function. Finally, note that  $w$  is surjective iff every letter in the alphabet  $A$  occurs among the letters  $w_1, \dots, w_k$ .

## 1.6 Bijections, Cardinality, and Counting

Bijections play a critical role in the theory of counting. Indeed, the very definition of cardinality is formulated in terms of bijections. In everyday life, we count the number of objects in a finite set  $S$  by pointing to each object in the set in turn and saying “one,” “two,” “three,” etc. In essence, we are setting up a bijection between  $S$  and some set  $\{1, 2, \dots, n\}$  of natural numbers. This leads to the following definition, which provides a rigorous foundation for the informal notion of cardinality that we have used up to this point.

**1.30. Definition: Cardinality.** For any set  $A$  and any integer  $n \geq 1$ , we write  $|A| = n$  iff there exists a bijection  $f : A \rightarrow \{1, 2, \dots, n\}$ . We write  $|A| = 0$  iff  $A = \emptyset$ . For any sets  $A$  and  $B$ , we write  $|A| = |B|$  iff there exists a bijection  $f : A \rightarrow B$ . We write  $|A| \leq |B|$  iff there exists an injection  $g : A \rightarrow B$ .

These definitions apply to infinite sets as well as finite sets, although we shall be mainly interested in finite sets. In the general case, one can prove the *Schröder-Bernstein Theorem*:  $|A| \leq |B|$  and  $|B| \leq |A|$  imply  $|A| = |B|$  (see [125, p. 29] or 1.156 for a proof.) If  $A$  is nonempty and the axiom of choice is assumed, then  $|A| \leq |B|$  is equivalent to the existence of a surjection  $h : B \rightarrow A$ . These properties are intuitively evident in the case of finite sets. For more discussion of the theory of cardinality for infinite sets, see [66] or [95].

Recall that if  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$  are functions, the *composition* of  $g$  and  $f$  is the function  $g \circ f : X \rightarrow Z$  defined by  $(g \circ f)(x) = g(f(x))$  for  $x \in X$ . We assume the reader is familiar with the following theorem, so we omit its proof.

**1.31. Theorem: Properties of Bijections.** Let  $X, Y, Z$  be any sets. (a) The identity map  $\text{id}_X : X \rightarrow X$ , defined by  $\text{id}_X(x) = x$  for all  $x \in X$ , is a bijection. Hence,  $|X| = |X|$ .



(b) A function  $f : X \rightarrow Y$  is bijective iff there exists a function  $f' : Y \rightarrow X$  such that  $f' \circ f = \text{id}_X$  and  $f \circ f' = \text{id}_Y$ . If such an  $f'$  exists, it is unique; we call it the *two-sided inverse* of  $f$  and denote it by  $f^{-1}$ . This inverse is also a bijection, and  $(f^{-1})^{-1} = f$ . Hence,  $|X| = |Y|$  implies  $|Y| = |X|$ . (c) The composition of two bijections is a bijection. Hence,  $|X| = |Y|$  and  $|Y| = |Z|$  implies  $|X| = |Z|$ .

The definition of cardinality can be used to *prove* the basic counting principle 1.1.

**1.32. Theorem.** If  $|A| = n$ ,  $|B| = m$ , and  $A \cap B = \emptyset$ , then  $|A \cup B| = n + m$ .

*Proof.* The assumption  $|A| = n$  means that there is a bijection  $f : A \rightarrow \{1, 2, \dots, n\}$ . The assumption  $|B| = m$  means that there is a bijection  $g : B \rightarrow \{1, 2, \dots, m\}$ . Define a function  $h : A \cup B \rightarrow \{1, 2, \dots, n + m\}$  by setting

$$h(x) = \begin{cases} f(x) & \text{if } x \in A; \\ g(x) + n & \text{if } x \in B. \end{cases}$$

The assumption that  $A \cap B = \emptyset$  is needed to ensure that  $h$  is a well-defined (single-valued) function. Observe that  $h$  does map into the required codomain  $\{1, 2, \dots, n + m\}$ . To see that  $h$  is a bijection, we display a two-sided inverse  $h' : \{1, 2, \dots, n + m\} \rightarrow A \cup B$ . We define

$$h'(i) = \begin{cases} f^{-1}(i) & \text{if } 1 \leq i \leq n; \\ g^{-1}(i - n) & \text{if } n + 1 \leq i \leq n + m. \end{cases}$$

A routine case analysis verifies that  $h \circ h'$  and  $h' \circ h$  are identity maps. □

The product rule 1.8 can be phrased more formally in terms of bijections.

**1.33. Formal Product Rule.** Suppose there is a bijection

$$f : \{1, 2, \dots, n_1\} \times \{1, 2, \dots, n_2\} \times \cdots \times \{1, 2, \dots, n_k\} \rightarrow S.$$

Then  $|S| = n_1 n_2 \cdots n_k$ .

*Proof.*  $S$  has the same cardinality as the product set  $\{1, 2, \dots, n_1\} \times \cdots \times \{1, 2, \dots, n_k\}$ , thanks to the bijection  $f$ . So the result follows from 1.5. □

**1.34. Remark.** Let us compare the formal product rule 1.33 to the informal version of the product rule given earlier (1.8). In informal applications of the product rule, we “build” objects in a set  $S$  by making a sequence of  $k$  choices, where there are  $n_i$  ways to make the  $i$ th choice. The input to the bijection  $f$  in the formal product rule is a  $k$ -tuple  $(c_1, \dots, c_k)$  where  $1 \leq c_i \leq n_i$  for all  $i \leq k$ . Intuitively,  $c_i$  records which choice was made at the  $i$ th stage. In practice, the map  $f$  is described as an algorithm that tells us how to combine the choices  $c_i$  to build an object in  $S$ . The key point in the intuitive product rule is that each object in  $S$  can be constructed *in exactly one way* by making suitable choices. This corresponds to the requirement in the formal product rule that  $f$  be a *bijection* onto  $S$ . Most erroneous applications of the intuitive product rule occur when the underlying “construction map”  $f$  is not bijective (a point that is seldom checked explicitly when using the product rule).

**1.35. Example.** How many 4-letter words contain at least one E? One might try to construct such words by choosing a position that contains the E (4 choices), then filling the remaining positions from left to right with arbitrary letters (26 choices for each position). The product rule would then give  $4 \times 26^3 = 70,304$  as the answer. However, this answer is incorrect. Our choice sequence implicitly defines a function

$$f : \{1, 2, 3, 4\} \times \{1, 2, \dots, 26\}^3 \rightarrow X,$$

where  $X$  is the set of words under consideration. For example,  $f(3, 3, 2, 26) = \text{CBEZ}$ . Our counting argument is flawed because the function  $f$  is surjective but not bijective. For instance,  $f(1, 1, 5, 1) = \text{EAEA} = f(3, 5, 1, 1)$ .

To obtain the correct answer, one can combine the product rule and the difference rule. There are  $26^4$  words of length 4, and there are  $25^4$  such words that do *not* contain the letter E. So the true answer is  $26^4 - 25^4 = 66,351$ . An alternative argument that is closer to our original attempt breaks  $X$  into the disjoint union  $X_1 \cup X_2 \cup X_3 \cup X_4$ , where  $X_i$  is the set of four-letter words where the *first* occurrence of E is at position  $i$ . A modification of the argument in the previous paragraph shows that  $|X_i| = 25^{i-1}26^{4-i}$ , so

$$|X| = 26^3 + 25 \cdot 26^2 + 25^2 \cdot 26 + 25^3 = 66,351.$$

**1.36. Remark.** One can give a *bijective* proof of the product rule (1.5), just as we gave a bijective proof of the sum rule (1.1) in 1.32. These bijective proofs have applications to the problems of listing, ranking, and unranking collections of combinatorial objects. These topics are discussed in Chapter 5.

## 1.7 Subsets, Binary Words, and Compositions

A fundamental method for counting a finite set  $A$  is to display a bijection between  $A$  and some other set  $B$  whose cardinality is already known. We illustrate this basic principle by revisiting the enumeration of subsets and binary words, and enumerating new combinatorial objects called compositions.

Let  $X = \{x, y, z\}$ , and consider the set  $\mathcal{P}(X)$  of all subsets of  $X$ . We define a bijection  $f : \mathcal{P}(X) \rightarrow \{0, 1\}^3$  as follows:

$$\begin{aligned} f(\emptyset) &= 000; & f(\{x\}) &= 100; & f(\{y\}) &= 010; & f(\{z\}) &= 001; \\ f(\{x, y\}) &= 110; & f(\{x, z\}) &= 101; & f(\{y, z\}) &= 011; & f(\{x, y, z\}) &= 111. \end{aligned}$$

These values were computed by the following rule. Given  $S \subseteq X$ , we set  $f(S) = w_1 w_2 w_3$  where  $w_1 = 1$  if  $x \in S$ ,  $w_1 = 0$  if  $x \notin S$ ,  $w_2 = 1$  if  $y \in S$ ,  $w_2 = 0$  if  $y \notin S$ ,  $w_3 = 1$  if  $z \in S$ , and  $w_3 = 0$  if  $z \notin S$ . We see by inspection that  $f$  is a bijection. Thus,  $|\mathcal{P}(X)| = |\{0, 1\}^3| = 2^3 = 8$ . We now generalize this example to  $n$ -element sets. First, we introduce notation that will be used frequently throughout the text.

**1.37. Definition: Truth Function.** If  $P$  is any logical statement, we set  $\chi(P) = 1$  if  $P$  is true, and  $\chi(P) = 0$  if  $P$  is false.

**1.38. Theorem: Subsets vs. Binary Words.** Let  $X$  be an  $n$ -element set. For each ordering  $x_1, \dots, x_n$  of the elements of  $X$ , there is a bijection  $f : \mathcal{P}(X) \rightarrow \{0, 1\}^n$ . Therefore,  $|\mathcal{P}(X)| = 2^n$ .

*Proof.* Given  $S \subseteq X$ , we define  $f(S) = w_1 w_2 \cdots w_n$ , where  $w_i = \chi(x_i \in S)$ . To see that  $f$  is a bijection, define  $f' : \{0, 1\}^n \rightarrow \mathcal{P}(X)$  by setting

$$f'(w_1 w_2 \cdots w_n) = \{x_i \in X : w_i = 1\}.$$

It is immediate that  $f'$  is the two-sided inverse of  $f$ . □

For example, if  $X = \{1, 2, 3, 4, 5, 6, 7, 8\}$  with the usual ordering, then  $f(\{2, 5, 7, 8\}) = 01001011$  and  $f^{-1}(10000011) = \{1, 7, 8\}$ .

**1.39. Definition: Compositions.** A *composition* of an integer  $n > 0$  is a sequence  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ , where each  $\alpha_i$  is a positive integer and  $\alpha_1 + \alpha_2 + \dots + \alpha_k = n$ . The *number of parts* of  $\alpha$  is  $k$ . Let  $\text{Comp}_n$  be the set of all compositions of  $n$ .

**1.40. Example.** The sequences  $(1, 3, 1, 3, 3)$  and  $(3, 3, 3, 1, 1)$  are two distinct compositions of 11 with five parts. The four compositions of 3 are

$$(3), \quad (2, 1), \quad (1, 2), \quad (1, 1, 1).$$

**1.41. Theorem: Enumeration of Compositions.** For all  $n > 0$ , there are  $2^{n-1}$  compositions of  $n$ .

*Proof.* We define a bijection  $g : \text{Comp}_n \rightarrow \{0, 1\}^{n-1}$ . Given  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k) \in \text{Comp}_n$ , define

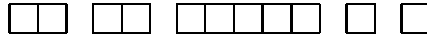
$$g(\alpha) = 0^{\alpha_1-1} 1 0^{\alpha_2-1} 1 \dots 1 0^{\alpha_k-1}.$$

Here, the notation  $0^j$  denotes a sequence of  $j$  consecutive zeroes, and  $0^0$  denotes the empty word. For example,  $g((3, 1, 3)) = 001100$ . Since  $\sum_{i=1}^k (\alpha_i - 1) = n - k$  and there are  $k - 1$  ones, we see that  $g(\alpha) \in \{0, 1\}^{n-1}$ . Now define  $g' : \{0, 1\}^{n-1} \rightarrow \text{Comp}_n$  as follows. We can uniquely write any word  $w \in \{0, 1\}^{n-1}$  in the form  $w = 0^{b_1} 1 0^{b_2} 1 \dots 1 0^{b_k}$  where  $k \geq 1$ , each  $b_i \geq 0$ , and  $\sum_{i=1}^k b_i = (n - 1) - (k - 1) = n - k$  since there are  $k - 1$  ones. Define  $g'(w) = (b_1 + 1, b_2 + 1, \dots, b_k + 1)$ , which is a composition of  $n$ . For example,  $g'(100100) = (1, 3, 3)$ . One may check that  $g'$  is the two-sided inverse of  $g$ , so  $g$  is a bijection. It follows that  $|\text{Comp}_n| = |\{0, 1\}^{n-1}| = 2^{n-1}$ .  $\square$

The bijections in the preceding proof are best understood pictorially. We represent an integer  $i > 0$  as a sequence of  $i$  unit squares glued together. We visualize a composition  $(\alpha_1, \dots, \alpha_k)$  by drawing the squares for  $\alpha_1, \dots, \alpha_k$  in a single row, separated by gaps. For instance, the composition  $(1, 3, 1, 3, 3)$  is represented by the picture



We now scan the picture from left to right and record what happens between each two successive boxes. If the two boxes in question are glued together, we record a 0; if there is a gap between the two boxes, we record a 1. The composition of 11 pictured above maps to the word  $1001100100 \in \{0, 1\}^{10}$ . Going the other way, the word  $0101000011 \in \{0, 1\}^{10}$  leads first to the picture



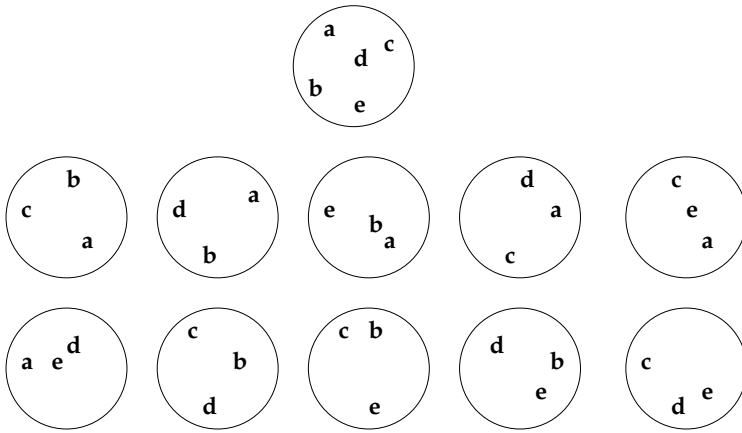
and then to the composition  $(2, 2, 5, 1, 1)$ . One can check that the pictorial operations just described correspond precisely to the maps  $f$  and  $f'$  in the proof above. When  $n = 3$ , we have:

$$f((3)) = 00; \quad f((2, 1)) = 01; \quad f((1, 2)) = 10; \quad f((1, 1, 1)) = 11.$$

## 1.8 Subsets of a Fixed Size

We turn now to the enumeration of the  $k$ -element subsets of an  $n$ -element set. For example, there are ten 3-element subsets of  $\{a, b, c, d, e\}$ :

$$\{a, b, c\}, \quad \{a, b, d\}, \quad \{a, b, e\}, \quad \{a, c, d\}, \quad \{a, c, e\},$$

**FIGURE 1.6**

The set  $\{a, b, c, d, e\}$  and its 3-element subsets.

$$\{a, d, e\}, \quad \{b, c, d\}, \quad \{b, c, e\}, \quad \{b, d, e\}, \quad \{c, d, e\}.$$

In this example, we present a given set by listing its members between curly braces. This notation forces us to list the members of each set in a particular order (alphabetical in this case). If we reorder the members of the list, the underlying set does not change. For example, the sets  $A_1 = \{a, c, d\}$  and  $A_2 = \{c, d, a\}$  and  $A_3 = \{d, c, a\}$  are all equal. This assertion follows from the very definition of set equality:  $A = B$  means that for every  $x$ ,  $x \in A$  iff  $x \in B$ . In contrast, the ordering of elements in a sequence (or word) definitely makes a difference. For instance, the words *cad* and *dac* are unequal although they use the same three letters.

To emphasize that the members of a set do not come in any particular order, we often picture a finite set as a circle with the members of the set floating around in random positions inside the circle. For example, Figure 1.6 depicts the sets mentioned above.

Suppose we try to enumerate the  $k$ -element subsets of a given  $n$ -element set using the product rule. Recall that the product rule requires us to construct objects by making an *ordered sequence* of choices. We might try to construct a subset by choosing its first element in  $n$  ways, then its second element in  $n - 1$  ways, etc., which leads to the *incorrect* answer  $n(n - 1) \cdots (n - k + 1)$ . The trouble here is that there is no well-defined “first element” of a subset. In fact, our naive construction procedure generates each subset several times, once for each possible ordering of its members. There are  $k!$  such orderings, so we obtain the correct answer by dividing the previous formula by  $k!$ . We make this argument more precise in the next theorem.

**1.42. Theorem: Enumeration of  $k$ -element Subsets.** For  $0 \leq k \leq n$ , the number of  $k$ -element subsets of an  $n$ -element set is

$$\frac{n!}{k!(n - k)!}.$$

*Proof.* Fix  $n$  and  $k$  with  $0 \leq k \leq n$ . Let  $A$  be an  $n$ -element set, and let  $x$  denote the number of  $k$ -element subsets of  $A$ . Let  $S$  be the set of all  $k$ -permutations of  $A$ . Recall that elements of  $S$  are *ordered* sequences  $w_1 w_2 \cdots w_k$ , where the  $w_i$  are distinct elements of  $A$ . We compute  $|S|$  in two ways. First, we have already seen that  $|S| = n!/(n - k)!$  by using the product rule — we choose  $w_1$  in  $n$  ways, then choose  $w_2$  in  $n - 1$  ways, etc., and finally

choose  $w_k$  in  $n - k + 1$  ways. On the other hand, here is a second way to construct a typical sequence  $w_1 w_2 \cdots w_k$  in  $S$ . Begin by choosing a  $k$ -element subset of  $A$  in any of  $x$  ways. Then write down a permutation of this  $k$ -element subset in any of  $k!$  ways. The result is an element of  $S$ . By the product rule,

$$x \cdot k! = |S| = n!/(n - k)!.$$

Solving for  $x$ , we obtain the desired formula.  $\square$

**1.43. Definition: Binomial Coefficients.** For  $0 \leq k \leq n$ , the *binomial coefficient* is

$$\binom{n}{k} = C(n, k) = \frac{n!}{k!(n - k)!}.$$

For  $k < 0$  or  $k > n$ , we define  $\binom{n}{k} = C(n, k) = 0$ . Thus, for all  $n \geq 0$  and all  $k$ ,  $\binom{n}{k}$  is the number of  $k$ -element subsets of an  $n$ -element set. In particular,  $\binom{n}{k}$  is always an integer.

## 1.9 Anagrams

**1.44. Definition: Anagrams.** Suppose  $a_1, \dots, a_k$  are distinct letters from some alphabet  $A$  and  $n_1, \dots, n_k$  are nonnegative integers. Let  $\mathcal{R}(a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k})$  denote the set of all words  $w = w_1 w_2 \cdots w_n$  that are formed by rearranging  $n_1$  copies of  $a_1$ ,  $n_2$  copies of  $a_2$ , ...,  $n_k$  copies of  $a_k$  (so that  $n = n_1 + n_2 + \cdots + n_k$ ). Words in a given set  $\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  are said to be *anagrams* or *rearrangements* of one another.

**1.45. Example.**

$$\mathcal{R}(0^2 1^3) = \{00111, 01011, 01101, 01110, 10011, 10101, 10110, 11001, 11010, 11100\};$$

$$\mathcal{R}(a^1 b^2 c^1 d^0) = \{abbc, abcb, acbb, abac, bacb, bbac, bbca, bcab, bcba, cabb, cbab, cbba\}.$$

**1.46. Theorem: Enumeration of Anagrams.** Suppose  $a_1, \dots, a_k$  are distinct letters,  $n_1, \dots, n_k$  are nonnegative integers, and  $n = n_1 + \cdots + n_k$ . Then

$$|\mathcal{R}(a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k})| = \frac{n!}{n_1! n_2! \cdots n_k!}.$$

*Proof.* We give two proofs of this result. *First Proof:* We use a technique similar to that used in the proof of 1.42. Define a new alphabet  $A$  consisting of  $n$  *distinct letters* by attaching distinct numerical superscripts to each copy of the given letters  $a_1, \dots, a_k$ :

$$A = \{a_1^{(1)}, a_1^{(2)}, \dots, a_1^{(n_1)}, a_2^{(1)}, \dots, a_2^{(n_2)}, \dots, a_k^{(1)}, \dots, a_k^{(n_k)}\}.$$

Let  $x = |\mathcal{R}(a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k})|$ . Let  $S$  be the set of all permutations  $w$  of  $A$ . We count  $|S|$  in two ways. On one hand, we already know that  $|S| = n!$  (choose  $w_1$  in  $n$  ways, then  $w_2$  in  $n - 1$  ways, etc.). On the other hand, here is a different method for constructing each permutation of  $A$  exactly once. First, choose a word  $v \in \mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  in any of  $x$  ways. Second, attach the superscripts 1 through  $n_1$  to the  $n_1$  copies of  $a_1$  in  $v$  in any of  $n_1!$  ways. Third, attach the superscripts 1 through  $n_2$  to the  $n_2$  copies of  $a_2$  in  $v$  in any of  $n_2!$  ways. Continue similarly; at the last stage, we attach the superscripts 1 through  $n_k$  to the  $n_k$  copies of  $a_k$  in  $v$  in any of  $n_k!$  ways. By the product rule,

$$x \cdot n_1! \cdot n_2! \cdot \dots \cdot n_k! = |S| = n!.$$

Solving for  $x$ , we obtain the desired formula.

*Second Proof.* The second proof relies on 1.42 and an algebraic manipulation of factorials. We construct a typical object  $w = w_1 w_2 \cdots w_n \in \mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  by making the following sequence of  $k$  choices. Intuitively, we are going to choose the positions of the  $a_1$ 's, then the positions of the  $a_2$ 's, etc. First, choose any  $n_1$ -element subset  $S_1$  of  $\{1, 2, \dots, n\}$  in any of  $\binom{n}{n_1}$  ways, and define  $w_i = a_1$  for all  $i \in S_1$ . Second, choose any  $n_2$ -element subset  $S_2$  of  $\{1, 2, \dots, n\} \sim S_1$  in any of  $\binom{n-n_1}{n_2}$  ways, and define  $w_i = a_2$  for all  $i \in S_2$ . At the  $j$ th stage (where  $1 \leq j \leq k$ ), we have already filled the positions in  $S_1 \cup \cdots \cup S_{j-1} \subseteq \{1, 2, \dots, n\}$ , and there are  $n - n_1 - n_2 - \cdots - n_{j-1}$  remaining positions in the word. We choose any  $n_j$ -element subset  $S_j$  of these remaining positions in any of  $\binom{n-n_1-\cdots-n_{j-1}}{n_j}$  ways, and define  $w_i = a_j$  for all  $i \in S_j$ . By the product rule, the number of rearrangements is

$$\binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \cdots \binom{n-n_1-\cdots-n_{k-1}}{n_k} = \prod_{i=1}^k \frac{(n-n_1-\cdots-n_{i-1})!}{n_i!(n-n_1-\cdots-n_i)!}.$$

This is a telescoping product that simplifies to  $n!/(n_1!n_2!\cdots n_k!)$ . For instance, when  $k = 4$ , the product is

$$\frac{n!}{n_1!(n-n_1)!} \cdot \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \cdot \frac{(n-n_1-n_2)!}{n_3!(n-n_1-n_2-n_3)!} \cdot \frac{(n-n_1-n_2-n_3)!}{n_4!(n-n_1-n_2-n_3-n_4)!},$$

which simplifies to  $n!/(n_1!n_2!n_3!n_4!)$ . (Recall that  $(n-n_1-n_2-\cdots-n_k)! = 0! = 1$ .)  $\square$

**1.47. Example.** We now illustrate the constructions in each of the two preceding proofs. For the first proof, suppose we are counting  $\mathcal{R}(a^3 b^1 c^4)$ . The alphabet  $A$  in the proof consists of the eight distinct letters

$$A = \{a^{(1)}, a^{(2)}, a^{(3)}, b^{(1)}, c^{(1)}, c^{(2)}, c^{(3)}, c^{(4)}\}.$$

Let us build a specific permutation of  $A$  using the second counting method. First, choose an element of  $\mathcal{R}(a^3 b^1 c^4)$ , say  $v = \text{baccaacc}$ . Second, choose a labeling of the  $a$ 's with superscripts, say  $ba^{(3)}cca^{(1)}a^{(2)}cc$ . Third, choose a labeling of the  $b$ 's, say  $b^{(1)}a^{(3)}cca^{(1)}a^{(2)}cc$ . Finally, choose a labeling of the  $c$ 's, say  $b^{(1)}a^{(3)}c^{(1)}c^{(2)}a^{(1)}a^{(2)}c^{(4)}c^{(3)}$ . We have now constructed a permutation of the alphabet  $A$ .

Next, let us see how to build the word 'baccaacc' using the method of the second proof. Start with an empty 8-letter word, which we denote ----- . We first choose the 3-element subset  $\{2, 5, 6\}$  of  $\{1, 2, \dots, 8\}$  and put  $a$ 's in those positions, obtaining -a--a a-- . We then choose the 1-element subset  $\{1\}$  of  $\{1, 3, 4, 7, 8\}$  and put a  $b$  in that position, obtaining b a--a a-- . Finally, we choose the 4-element subset  $\{3, 4, 7, 8\}$  of  $\{3, 4, 7, 8\}$  and put  $c$ 's in those positions, obtaining the word baccaacc.

**1.48. Definition: Multinomial Coefficients.** Suppose  $n_1, \dots, n_k$  are nonnegative integers and  $n = n_1 + \cdots + n_k$ . The *multinomial coefficient* is

$$\binom{n}{n_1, n_2, \dots, n_k} = C(n; n_1, n_2, \dots, n_k) = \frac{n!}{n_1!n_2!\cdots n_k!}.$$

This is the number of rearrangements of  $k$  letters where there are  $n_i$  copies of the  $i$ th letter. In particular,  $\binom{n}{n_1, n_2, \dots, n_k}$  is always an integer.

**1.49. Theorem: Binomial vs. Multinomial Coefficients.** For all nonnegative integers  $a$  and  $b$ , we have

$$\binom{a+b}{a} = \binom{a+b}{a, b}.$$

*Proof.* The result is immediate from the formulas for binomial coefficients and multinomial coefficients as quotients of factorials, but we want to give a bijective proof. Let  $U$  be the set of  $a$ -element subsets of  $\{1, 2, \dots, a+b\}$ , and let  $V = \mathcal{R}(1^a 0^b)$ . We have already shown that  $|U| = \binom{a+b}{a}$  and  $|V| = \binom{a+b}{a,b}$ . So we must define a bijection  $f : U \rightarrow V$ . Given  $S \in U$ , let  $f(S) = w_1 w_2 \dots w_{a+b}$ , where  $w_i = \chi(i \in S)$ . Since  $S$  has  $a$  elements,  $f(S)$  is a word consisting of  $a$  ones and  $b$  zeroes. The inverse of  $f$  is the map  $f' : V \rightarrow U$  given by  $f'(w_1 w_2 \dots w_{a+b}) = \{i : w_i = 1\}$ . (Note that these maps are the restrictions to  $U$  and  $V$  of the maps  $f$  and  $f'$  from the proof of 1.38.)  $\square$

**1.50. Example: Compositions with  $k$  Parts.** Let us determine the number of compositions  $\alpha = (\alpha_1, \dots, \alpha_k)$  of  $n$  that have exactly  $k$  parts. Recall the bijection  $g : \text{Comp}_n \rightarrow \{0, 1\}^{n-1}$  from the proof of 1.41. Applying  $g$  to  $\alpha$  produces a word with  $k-1$  ones and  $n-k$  zeroes. Conversely, any such word arises from a composition with  $k$  parts. Thus,  $g$  restricts to a bijection between the set of compositions of  $n$  with  $k$  parts and the set of words  $\mathcal{R}(0^{n-k} 1^{k-1})$ . Consequently, the number of such compositions is  $\binom{n-1}{n-k, k-1}$ .

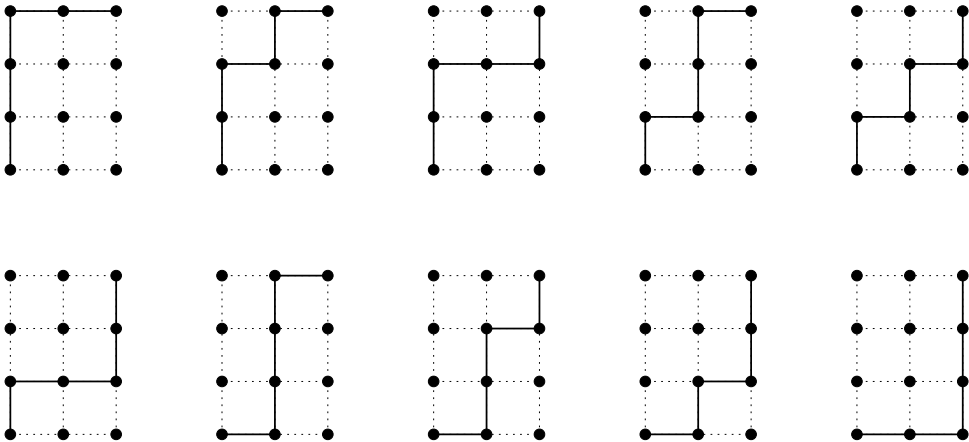
## 1.10 Lattice Paths

**1.51. Definition: Lattice Paths.** A *lattice path* in the plane is a sequence

$$P = ((x_0, y_0), (x_1, y_1), \dots, (x_k, y_k)),$$

where the  $x_i$ 's and  $y_i$ 's are integers, and for each  $i \geq 1$ , either  $(x_i, y_i) = (x_{i-1} + 1, y_{i-1})$  or  $(x_i, y_i) = (x_{i-1}, y_{i-1} + 1)$ . We say that  $P$  is a path *from*  $(x_0, y_0)$  *to*  $(x_k, y_k)$ .

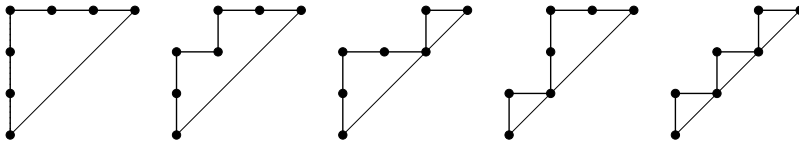
We often take  $(x_0, y_0)$  to be the origin  $(0, 0)$ . We represent  $P$  pictorially by drawing a line segment of length 1 from  $(x_{i-1}, y_{i-1})$  to  $(x_i, y_i)$  for each  $i$ . For example, Figure 1.7 displays the ten lattice paths from  $(0, 0)$  to  $(2, 3)$ .



**FIGURE 1.7**

Lattice paths from  $(0, 0)$  to  $(2, 3)$ .

**1.52. Theorem: Enumeration of Lattice Paths in a Rectangle.** For all integers  $a, b \geq 0$ , there are  $\binom{a+b}{a,b} = \frac{(a+b)!}{a!b!}$  lattice paths from  $(0, 0)$  to  $(a, b)$ .

**FIGURE 1.8**

Dyck paths of order 3.

*Proof.* We can encode a lattice path  $P$  from  $(0,0)$  to  $(a,b)$  as a word  $w \in \mathcal{R}(E^a N^b)$  by setting  $w_i = E$  if  $(x_i, y_i) = (x_{i-1} + 1, y_{i-1})$  and  $w_i = N$  if  $(x_i, y_i) = (x_{i-1}, y_{i-1} + 1)$ . Here,  $E$  stands for “east step,” and  $N$  stands for “north step.” Since the path ends at  $(a,b)$ ,  $w$  must have exactly  $a$  occurrences of  $E$  and exactly  $b$  occurrences of  $N$ . Thus we have a bijection between the given set of lattice paths and the set  $\mathcal{R}(E^a N^b)$ . Since  $|\mathcal{R}(E^a N^b)| = \binom{a+b}{a,b}$ , the theorem follows.  $\square$

For example, the paths shown in Figure 1.7 are encoded by the words

$$\begin{array}{cccccc} \text{NNNEE,} & \text{NNENE,} & \text{NNEEN,} & \text{NENNE,} & \text{NENEN,} \\ \text{NEENN,} & \text{ENNNE,} & \text{ENNEN,} & \text{ENENN,} & \text{EENNN.} \end{array}$$

More generally, one can consider lattice paths in  $\mathbb{R}^d$ . Such a path is a sequence of points  $(v_0, v_1, \dots, v_k)$  in  $\mathbb{Z}^d$  such that for each  $i$ ,  $v_i = v_{i-1} + e_j$  for some standard basis vector  $e_j = (0, \dots, 1, \dots, 0) \in \mathbb{R}^d$  (the 1 occurs in position  $j$ ).

**1.53. Theorem: Enumeration of Lattice Paths in a  $d$ -dimensional Rectangle.** For all integers  $n_1, \dots, n_d \geq 0$ , the number of  $d$ -dimensional lattice paths from  $(0, \dots, 0)$  to  $(n_1, \dots, n_d)$  is

$$|\mathcal{R}(e_1^{n_1} e_2^{n_2} \dots e_d^{n_d})| = \binom{n_1 + n_2 + \dots + n_d}{n_1, n_2, \dots, n_d}.$$

*Proof.* Encode a path  $P$  by the word  $w_1 w_2 \dots w_n$ , where  $n = n_1 + \dots + n_d$  and  $w_i = e_j$  iff  $v_i = v_{i-1} + e_j$ .  $\square$

Henceforth, we usually will make no distinction between a lattice path (which is a sequence of lattice points) and the word that encodes the lattice path.

We now turn to a more difficult enumeration problem involving lattice paths.

**1.54. Definition: Dyck Paths.** A *Dyck path of order  $n$*  is a lattice path from  $(0,0)$  to  $(n,n)$  such that  $y_i \geq x_i$  for all points  $(x_i, y_i)$  on the path. This requirement means that the path always stays weakly above the line  $y = x$ . For example, Figure 1.8 displays the five Dyck paths of order 3.

**1.55. Definition: Catalan Numbers.** For  $n \geq 0$ , the  $n$ th *Catalan number* is

$$C_n = \frac{1}{n+1} \binom{2n}{n} = \frac{1}{2n+1} \binom{2n+1}{n+1, n} = \frac{(2n)!}{n!(n+1)!} = \binom{2n}{n, n} - \binom{2n}{n+1, n-1}.$$

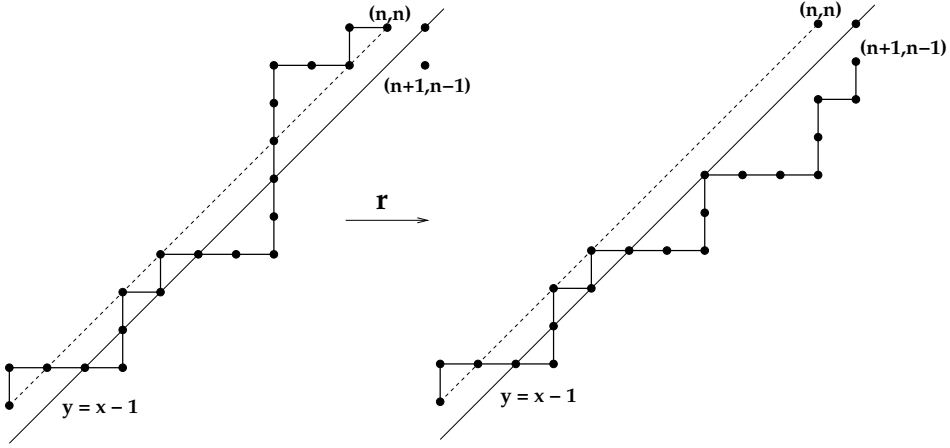
One may check that these expressions are all equal. For instance,

$$\binom{2n}{n, n} - \binom{2n}{n+1, n-1} = \frac{(2n)!}{n!n!} - \frac{(2n)!}{(n+1)!(n-1)!} = \frac{(2n)!}{n!n!} \left[ 1 - \frac{n}{n+1} \right] = \frac{1}{n+1} \binom{2n}{n, n}.$$

The first few Catalan numbers are

$$C_0 = 1, \quad C_1 = 1, \quad C_2 = 2, \quad C_3 = 5, \quad C_4 = 14, \quad C_5 = 42, \quad C_6 = 132, \quad C_7 = 429.$$



**FIGURE 1.9**

Example of the reflection map  $r$ .

**1.56. Theorem: Enumeration of Dyck Paths.** For  $n \geq 0$ , the number of Dyck paths of order  $n$  is the Catalan number  $C_n = \binom{2n}{n,n} - \binom{2n}{n+1,n-1}$ .

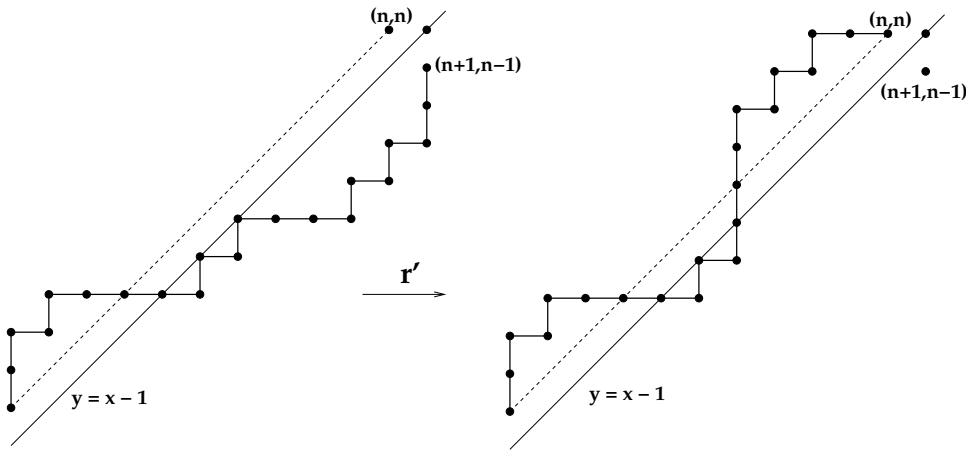
*Proof.* The following proof is essentially due to André [3]. Let  $A$  be the set of all lattice paths from  $(0,0)$  to  $(n,n)$ ; let  $B$  be the set of all lattice paths from  $(0,0)$  to  $(n+1, n-1)$ ; let  $C$  be the set of all Dyck paths of order  $n$ ; and let  $D = A \sim C$  be the set of paths from  $(0,0)$  to  $(n,n)$  that do go strictly below the line  $y = x$ . Since  $C = A \sim D$ , the difference rule gives

$$|C| = |A| - |D|.$$

We already know that  $|A| = \binom{2n}{n,n}$  and  $|B| = \binom{2n}{n+1,n-1}$ . To establish the desired formula  $|C| = C_n$ , it therefore suffices to exhibit a bijection  $r : D \rightarrow B$ .

We define  $r$  as follows. Given a path  $P \in D$ , follow the path backwards from  $(n,n)$  until it goes below the diagonal  $y = x$  for the first time. Let  $(x_i, y_i)$  be the first lattice point we encounter that is below  $y = x$ ; this point must lie on the line  $y = x - 1$ .  $P$  is the concatenation of two lattice paths  $P_1$  and  $P_2$ , where  $P_1$  goes from  $(0,0)$  to  $(x_i, y_i)$  and  $P_2$  goes from  $(x_i, y_i)$  to  $(n,n)$ . By choice of  $i$ , every lattice point of  $P_2$  after  $(x_i, y_i)$  lies strictly above the line  $y = x - 1$ . Now, let  $P'_2$  be the path from  $(x_i, y_i)$  to  $(n+1, n-1)$  obtained by reflecting  $P_2$  in the line  $y = x - 1$ . Define  $r(P)$  to be the concatenation of  $P_1$  and  $P'_2$ . See Figure 1.9 for an example. Here,  $(x_i, y_i) = (7, 6)$ ,  $P_1 = \text{NEEENNNEEEENN}$ ,  $P_2 = \text{NNNEENE}$ , and  $P'_2 = \text{EEENNEN}$ . Note that  $r(P)$  is a lattice path from  $(0,0)$  to  $(n+1, n-1)$ , so  $r(P) \in B$ . Furthermore,  $(x_i, y_i)$  is the only lattice point of  $P'_2$  lying on the line  $y = x - 1$ .

The inverse map  $r' : B \rightarrow D$  acts as follows. Given  $Q \in B$ , choose  $i$  maximal such that  $(x_i, y_i)$  is a point of  $Q$  on the line  $y = x - 1$ . Such an  $i$  must exist, since there is no way for a lattice path to reach  $(n+1, n-1)$  from  $(0,0)$  without passing through this line. Write  $Q = Q_1Q_2$ , where  $Q_1$  goes from  $(0,0)$  to  $(x_i, y_i)$  and  $Q_2$  goes from  $(x_i, y_i)$  to  $(n+1, n-1)$ . Let  $Q'_2$  be the reflection of  $Q_2$  in the line  $y = x - 1$ . Define  $r'(Q) = Q_1Q'_2$ , and note that this is a lattice path from  $(0,0)$  to  $(n,n)$  which passes through  $(x_i, y_i)$ , and hence lies in  $D$ . See Figure 1.10 for an example. Here,  $(x_i, y_i) = (6, 5)$ ,  $Q_1 = \text{NNNEEEEEENEN}$ ,  $Q_2 = \text{EEENENENN}$ , and  $Q'_2 = \text{NNNENENEE}$ . From our observations about the point  $(x_i, y_i)$  in this paragraph and the last, one sees that  $r'$  is the two-sided inverse of  $r$ .  $\square$

**FIGURE 1.10**

Example of the inverse reflection map.

The technique used in the preceding proof is called *André's reflection principle*. Another proof of the theorem, which leads directly to the formula  $\frac{1}{2n+1} \binom{2n+1}{n+1, n}$ , is given in §12.1. Yet another proof, which leads directly to the formula  $\frac{1}{n+1} \binom{2n}{n, n}$ , is given in §12.2.

## 1.11 Multisets

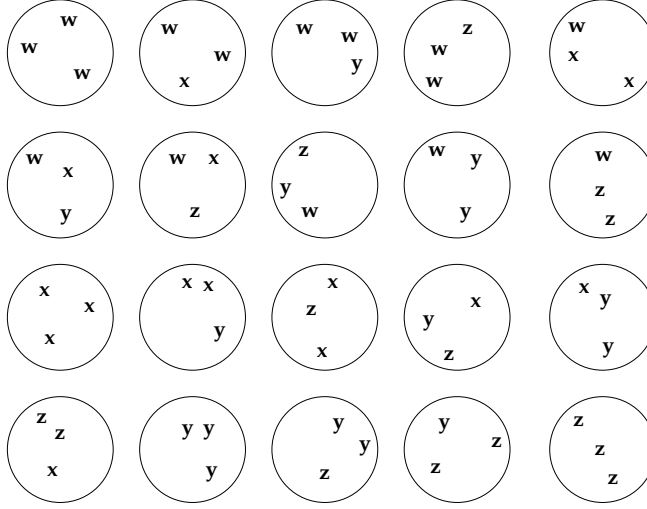
Recall that the concepts of *order* and *multiplicity* play no role when deciding whether two *sets* are equal. For instance,  $\{1, 3, 5\} = \{3, 5, 1\} = \{1, 1, 1, 5, 5, 3, 3, 3\}$  since all these sets have the same members. We now introduce the concept of a *multiset*, in which order still does not matter, but repetitions of a given element are significant.

**1.57. Definition: Multisets.** A *multiset* is an ordered pair  $M = (S, m)$ , where  $S$  is a set and  $m : S \rightarrow \mathbb{N}^+$  is a function. For  $x \in S$ , the number  $m(x)$  is called the *multiplicity* of  $x$  in  $M$ . The *number of elements* of  $M$  is  $|M| = \sum_{x \in S} m(x)$ .

In contrast, the number of *distinct* elements in  $M$  is  $|S|$ . We sometimes display a multiset as a list  $[x_1, x_2, \dots, x_k]$ , where each  $x \in S$  occurs exactly  $m(x)$  times in the list. However, one must remember that the order of the elements in this list does not matter when deciding equality of multisets. For example,  $[1, 1, 2, 3, 3] \neq [1, 1, 1, 2, 3, 3] = [3, 2, 3, 1, 1, 1]$ . We often visualize  $M$  as a circle with the elements  $x \in S$  appearing inside, each repeated  $m(x)$  times. For example, Figure 1.11 displays the twenty 3-element multisets using letters in the alphabet  $\{w, x, y, z\}$ . The last circle in the first row represents the multiset  $[w, x, x]$ , which is formally the ordered pair  $(\{w, x\}, m)$  such that  $m(w) = 1$  and  $m(x) = 2$ .

**1.58. Theorem: Enumeration of Multisets.** The number of  $k$ -element multisets using letters from an  $n$ -letter alphabet is

$$\binom{k+n-1}{k, n-1} = \frac{(k+n-1)!}{k!(n-1)!}.$$

**FIGURE 1.11**

The 3-element multisets over the alphabet  $\{w, x, y, z\}$ .

*Proof.* We give two proofs of this result. *First Proof:* Let  $A$  be a fixed  $n$ -letter alphabet, and let  $U$  be the set of all  $k$ -element multisets using letters from  $A$ . Introduce the two symbols  $\star$  (“star”) and  $|$  (“bar”), and let  $V = \mathcal{R}(\star^k |^{n-1})$  be the set of all rearrangements of  $k$  stars and  $n - 1$  bars. We know that  $|V| = \binom{k+n-1}{k, n-1}$ . It therefore suffices to define a bijection  $f : U \rightarrow V$ .

Let  $(a_1, a_2, \dots, a_n)$  be a fixed ordering of the alphabet  $A$ . Let  $M = (S, m)$  be a typical multiset in  $U$ . Set  $m(a_i) = 0$  if  $a_i \notin S$ . Define

$$f(M) = \star^{m(a_1)} | \star^{m(a_2)} | \dots | \star^{m(a_{n-1})} | \star^{m(a_n)} \in V.$$

In other words, we write a star for each occurrence of  $a_1$  in  $M$  (if any), then a bar, then a star for each occurrence of  $a_2$  in  $M$  (if any), then a bar, etc. There is no bar after the stars for  $a_n$ , so there are only  $n - 1$  bars total. Since  $M$  has  $k$  elements, there are  $k$  stars total. Thus  $f(M)$  really is an element of  $V$ . For example, the multisets in the first column of Figure 1.11 are mapped to the following star-bar words:

$$f([w, w, w]) = \star\star\star||, \quad f([w, x, y]) = \star|\star|\star|, \quad f([x, x, x]) = |\star\star\star|, \quad f([x, z, z]) = |\star||\star\star.$$

The multiset  $M$  is uniquely determined by  $f(M)$ . More precisely, define  $f' : V \rightarrow U$  by letting  $f'(\star^{m_1} | \star^{m_2} | \dots | \star^{m_n})$  be the unique multiset that has  $m_i$  copies of  $a_i$  for  $1 \leq i \leq n$  (here  $m_i \geq 0$ ). Since  $\sum_{i=1}^n m_i = k$ , this is a  $k$ -element multiset using letters from  $A$ . For example, if  $n = 6$ ,  $k = 4$ , and  $A = \{1, 2, 3, 4, 5, 6\}$ , then  $f'(|\star||\star|\star|\star) = [3, 5, 6, 6]$ . One may check that  $f'$  is the two-sided inverse of  $f$ .

*Second Proof:* We may assume (without loss of generality) that the alphabet  $A$  is  $\{1, 2, \dots, n\}$ . As above, let  $U$  be the set of all  $k$ -element multisets using letters from  $A$ . Let  $W$  be the set of all  $k$ -element subsets of  $B = \{1, 2, \dots, k + n - 1\}$ . We know that  $|W| = \binom{k+n-1}{k} = \binom{k+n-1}{k, n-1}$ . So it suffices to define a bijection  $g : U \rightarrow W$ .

Given  $M \in U$ , we can write  $M$  uniquely in the form  $M = [x_1, x_2, \dots, x_k]$  by requiring that  $x_1 \leq x_2 \leq \dots \leq x_k$ . Now define

$$g(M) = g([x_1, x_2, \dots, x_k]) = \{x_1 + 0, x_2 + 1, x_3 + 2, \dots, x_i + (i - 1), \dots, x_k + (k - 1)\}.$$

For example, if  $n = 5$ ,  $k = 5$ , and  $M = [1, 1, 4, 5, 5]$ , then  $g(M) = \{1, 2, 6, 8, 9\} \subseteq \{1, 2, \dots, 9\}$ . Notice that the elements of the set  $g(M)$  all lie in  $\{1, 2, \dots, k + n - 1\}$  since  $1 \leq x_i \leq n$  for all  $i$ . Also, the  $k$  displayed elements of  $g(M)$  are pairwise distinct because, for any  $i < j$ , the assumption  $x_i \leq x_j$  implies  $x_i + (i - 1) < x_j + (j - 1)$ . Thus,  $g(M)$  is indeed a  $k$ -element subset of  $B$ .

Going the other way, define  $g' : W \rightarrow U$  as follows. Given  $S \in W$ , we can write  $S$  uniquely in the form  $S = \{y_1, y_2, \dots, y_k\}$  where  $y_1 < y_2 < \dots < y_k$ . Now define

$$g'(S) = g'(\{y_1, y_2, \dots, y_k\}) = [y_1 - 0, y_2 - 1, \dots, y_i - (i - 1), \dots, y_k - (k - 1)].$$

For example, if  $n = k = 5$  and  $S = \{2, 3, 5, 7, 8\}$ , then  $g'(S) = [2, 2, 3, 4, 4] \in U$ . Since every  $y_j \geq 1$  and the  $y_i$ 's form a strictly increasing sequence of integers, it follows that  $i \leq y_i$  for all  $i$ . Similarly, since every  $y_j \leq k + n - 1$  and there are  $k - i$  entries that exceed  $y_i$  in the sequence (namely  $y_{i+1}, \dots, y_k$ ), we deduce that  $y_i \leq (k + n - 1) - (k - i) = n + i - 1$  for all  $i$ . Subtracting  $i - 1$ , it follows that every element of the  $k$ -element multiset  $g'(S)$  lies in the range  $\{1, 2, \dots, n\}$ , so that  $g'(S)$  really is an element of  $U$ . It is now routine to check that  $g'$  is the two-sided inverse of  $g$ .  $\square$

## 1.12 Probability

The basic techniques of counting can be applied to solve a number of problems from probability theory. This section introduces some fundamental concepts of probability and considers several examples.

**1.59. Definition: Sample Spaces and Events.** A *sample space* is a set  $S$ , whose members represent the possible outcomes of a “random experiment.” In this section, we only consider *finite* sample spaces. An *event* is a subset of the sample space.

Intuitively, an event consists of the set of outcomes of the experiment that possess a particular property we are interested in.

**1.60. Example: Coin Tossing.** Suppose the experiment consists of tossing a coin five times. We could take the sample space for this experiment to be  $S = \{H, T\}^5$ , the set of all 5-letter words using the letters  $H$  (for heads) and  $T$  (for tails). The element  $\text{HHH}T\text{H} \in S$  represents the outcome where the fourth toss was tails and all other tosses were heads. The subset  $A = \{w \in S : w_1 = H\}$  is the event in which the first toss comes up heads. The subset  $B = \{w \in S : w_1 \neq w_5\}$  is the event that the first toss is different from the last toss. The subset

$$C = \{w \in S : w_i = T \text{ for an odd number of indices } i\}$$

is the event that we get an odd number of tails.

**1.61. Example: Dice Rolling.** Suppose the experiment consists of rolling a six-sided die three times. The sample space for this experiment is  $S = \{1, 2, 3, 4, 5, 6\}^3$ , the set of all 3-letter words over the alphabet  $\{1, 2, \dots, 6\}$ . The subset  $A = \{w \in S : w_1 + w_2 + w_3 \in \{7, 11\}\}$  is the event that the sum of the three numbers rolled is 7 or 11. The subset  $B = \{w \in S : w_1 = w_2 = w_3\}$  is the event that all three numbers rolled are the same. The subset  $C = \{w \in S : w \neq (4, 1, 3)\}$  is the event that we do not see the numbers 4, 1, 3 (in that order) in the dice rolls.

**1.62. Example: Lotteries.** Consider the following random experiment. We put 49 white balls (numbered 1 through 49) into a machine that mixes the balls for awhile and then outputs a sequence of six distinct balls, one at a time. We could take the sample space here to be the set  $S'$  of all 6-letter words  $w$  consisting of six distinct letters from  $A = \{1, 2, \dots, 49\}$ . In lotteries, the order in which the balls are drawn usually does not matter, so it is more common to take the sample space to be the set  $S$  of all 6-element subsets of  $A$ . (We will see later that using  $S$  instead of  $S'$  does not affect the probabilities we are interested in.) Suppose a lottery player picks a (fixed and known) 6-element subset  $T_0$  of  $A$ . For  $0 \leq k \leq 6$ , define events  $B_k = \{T \in S : |T \cap T_0| = k\} \subseteq S$ . Intuitively, the event  $B_k$  is the set of outcomes in which the player has matched exactly  $k$  of the winning lottery numbers.

**1.63. Example: Special Events.** For any sample space  $S$ ,  $\emptyset$  and  $S$  are events. Intuitively, the event  $\emptyset$  contains no outcomes, and therefore “never happens.” On the other hand, the event  $S$  contains all the outcomes, and therefore “always happens.” If  $A$  and  $B$  are *events* (i.e., subsets of  $S$ ), note that  $A \cup B$ ,  $A \cap B$ ,  $S \sim A$ , and  $A \sim B$  are also *events*. Intuitively,  $A \cup B$  is the event that either  $A$  happens or  $B$  happens (or both);  $A \cap B$  is the event that both  $A$  and  $B$  happen;  $S \sim A$  is the event that  $A$  does not happen; and  $A \sim B$  is the event that  $A$  happens but  $B$  does not happen.

Now we can formally define the concept of probability. Intuitively, for each event  $A$ , we want to define a number  $P(A)$  that measures the probability or likelihood that  $A$  occurs. Numbers close to 1 represent more likely events, while numbers close to 0 represent less likely events. A probability-zero event is “impossible,” while a probability-one event is “certain” to occur.

**1.64. Definition: Probability.** Assume  $S$  is a finite sample space. Recall that  $\mathcal{P}(S)$  is the set of all subsets of  $S$ , i.e., the set of all events. A *probability measure* for  $S$  is a function  $P : \mathcal{P}(S) \rightarrow [0, 1]$  such that  $P(\emptyset) = 0$ ;  $P(S) = 1$ ; and for any two *disjoint* events  $A$  and  $B$ ,  $P(A \cup B) = P(A) + P(B)$ .

By induction, it follows that  $P$  satisfies the *finite additivity* property

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

for all pairwise disjoint sets  $A_1, A_2, \dots, A_n \subseteq S$ .

**1.65. Example: Classical Probability Spaces.** Suppose  $S$  is a finite sample space in which all outcomes are equally likely. Then we must have  $P(\{x\}) = 1/|S|$  for each outcome  $x \in S$ . For any event  $A \subseteq S$ , finite additivity gives

$$P(A) = \frac{|A|}{|S|} = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}}. \quad (1.1)$$

Thus the calculation of probabilities (in this classical setup) reduces to two counting problems: counting the number of elements in  $A$  and counting the number of elements in  $S$ . We can take equation (1.1) as the *definition* of our probability measure  $P$ . Note that the axiom  $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$  is then a consequence of the sum rule. Also note that this probability model will only be appropriate if all the possible outcomes of the underlying random experiment are equally likely to occur.

**1.66. Example: Coin Tossing.** Suppose we toss a fair coin five times. The sample space is  $S = \{H, T\}^5$ , so that  $|S| = 2^5 = 32$ . Consider the event  $A = \{w \in S : w_1 = H\}$  of getting a head on the first toss. By the product rule,  $|A| = 1 \cdot 2^4 = 16$ , so  $P(A) = 16/32 = 1/2$ . Consider the event  $B = \{w \in S : w_1 \neq w_5\}$  in which the first toss differs from the last toss.

$B$  is the disjoint union of  $B_1 = \{w \in S : w_1 = H, w_5 = T\}$  and  $B_2 = \{w \in S : w_1 = T, w_5 = H\}$ . The product rule shows that  $|B_1| = |B_2| = 2^3 = 8$ , so that  $P(B) = (8 + 8)/32 = 1/2$ . Finally, consider the event

$$C = \{w \in S : w_i = T \text{ for an odd number of indices } i\}.$$

$C$  is the disjoint union  $C_1 \cup C_3 \cup C_5$ , where (for  $0 \leq k \leq 5$ )  $C_k$  is the event of getting exactly  $k$  tails. We have  $C_k = \mathcal{R}(T^k H^{5-k})$ , so that  $P(C_k) = \binom{5}{k}/2^5$ . Therefore,

$$P(C) = \frac{\binom{5}{1} + \binom{5}{3} + \binom{5}{5}}{2^5} = 16/32 = 1/2.$$

**1.67. Example: Dice Rolling.** Consider the experiment of rolling a six-sided die twice. The sample space is  $S = \{1, 2, 3, 4, 5, 6\}^2$ , so that  $|S| = 6^2 = 36$ . Consider the event  $A = \{x \in S : x_1 + x_2 \in \{7, 11\}\}$  of rolling a sum of 7 or 11. By direct enumeration, we have

$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1), (5, 6), (6, 5)\}; \quad |A| = 8.$$

Therefore,  $P(A) = 8/36 = 2/9$ . Consider the event  $B = \{x \in S : x_1 \neq x_2\}$  of getting two different numbers on the two rolls. The product rule gives  $|B| = 6 \cdot 5 = 30$ , so  $P(B) = 30/36 = 5/6$ .

**1.68. Example: Balls in Urns.** Suppose an urn contains  $n_1$  red balls,  $n_2$  white balls, and  $n_3$  blue balls. Let the random experiment consist of randomly drawing a  $k$ -element subset of balls from the urn. What is the probability of drawing  $k_1$  red balls,  $k_2$  white balls, and  $k_3$  blue balls, where  $k_1 + k_2 + k_3 = k$ ? We can take the sample space  $S$  to be all  $k$ -element subsets of the set

$$\{1, 2, \dots, n_1, n_1 + 1, \dots, n_1 + n_2, n_1 + n_2 + 1, \dots, n_1 + n_2 + n_3\}.$$

Here the first  $n_1$  integers represent red balls, the next  $n_2$  integers represent white balls, and the last  $n_3$  integers represent blue balls. We know that  $|S| = \binom{n_1 + n_2 + n_3}{k}$ . Let  $A$  be the event where we draw  $k_1$  red balls,  $k_2$  white balls, and  $k_3$  blue balls. To build a set  $T \in A$ , we choose a  $k_1$ -element subset of  $\{1, 2, \dots, n_1\}$ , then a  $k_2$ -element subset of  $\{n_1 + 1, \dots, n_1 + n_2\}$ , then a  $k_3$ -element subset of  $\{n_1 + n_2 + 1, \dots, n_1 + n_2 + n_3\}$ . By the product rule,  $|A| = \binom{n_1}{k_1} \binom{n_2}{k_2} \binom{n_3}{k_3}$ . Therefore, the definition of the probability measure gives

$$P(A) = \frac{\binom{n_1}{k_1} \binom{n_2}{k_2} \binom{n_3}{k_3}}{\binom{n_1 + n_2 + n_3}{k_1 + k_2 + k_3}}.$$

This calculation can be generalized to the case where the urn has balls of more than three colors.

**1.69. Example: Lotteries.** Consider the lottery described in 1.62. Here the sample space  $S$  consists of all 6-element subsets of  $A = \{1, 2, \dots, 49\}$ , so  $|S| = \binom{49}{6} = 13,983,816$ . Suppose a lottery player picks a (fixed and known) 6-element subset  $T_0$  of  $A$ . For  $0 \leq k \leq 6$ , define events  $B_k = \{T \in S : |T \cap T_0| = k\}$ .  $B_k$  occurs when the player matches exactly  $k$  of the winning numbers. We can build a typical object  $T \in B_k$  by choosing  $k$  elements of  $T_0$  in  $\binom{6}{k}$  ways, and then choosing  $6 - k$  elements of  $A \sim T_0$  in  $\binom{43}{6-k}$  ways. Hence,

$$P(B_k) = \frac{\binom{6}{k} \binom{43}{6-k}}{\binom{49}{6}}.$$

**TABLE 1.2**

Analysis of Virginia's "Lotto South" lottery.

Matches	Probability	Prize Value
3	0.01765 or 1 in 57	about \$5
4	0.0009686 or 1 in 1032	about \$75
5	0.00001845 or 1 in 54,201	about \$1000
6	$7.15 \times 10^{-8}$ or 1 in 13,983,816	jackpot

Table 1.2 shows the probability of matching  $k$  numbers, for  $3 \leq k \leq 6$ . The table also shows the amount of money one would win in the various cases. One can view this example as the special case of the previous example where the urn contains 6 balls of one color and 43 balls of another color.

In the lottery example, suppose we took the sample space to be the set  $S'$  of all *ordered* sequences of six distinct elements of  $\{1, 2, \dots, 49\}$ . Let  $B'_k$  be the event that the player guesses exactly  $k$  numbers correctly (disregarding order, as usual). Let  $P'$  be the probability measure on the sample space  $S'$ . One may check that  $|S'| = \binom{49}{6} \cdot 6!$  and  $|B'_k| = \binom{6}{k} \binom{43}{6-k} \cdot 6!$ , so that

$$P'(B'_k) = \frac{\binom{6}{k} \binom{43}{6-k} 6!}{\binom{49}{6} 6!} = P(B_k).$$

This confirms our earlier remark that the two sample spaces  $S$  and  $S'$  give the same probabilities for events that do not depend on the order in which the balls are drawn.

**1.70. Example: Lattice Paths.** Suppose we randomly choose a lattice path from  $(0, 0)$  to  $(n, n)$ . What is the probability that this path is a Dyck path? We know that there are  $\frac{1}{n+1} \binom{2n}{n}$  Dyck paths and  $\binom{2n}{n}$  lattice paths ending at  $(n, n)$ . Therefore, the probability is  $1/(n+1)$ . We discuss a remarkable generalization of this result, called the *Chung-Feller Theorem*, in §12.2.

**1.71. Example: General Probability Measures on a Finite Sample Space.** We now extend the previous discussion to the case where not all outcomes of the random experiment are equally likely. Let  $S$  be a finite sample space and let  $p : S \rightarrow [0, 1]$  be a map such that  $\sum_{x \in S} p(x) = 1$ . Intuitively,  $p(x)$  is the probability that the outcome  $x$  occurs. Now  $p$  is not a probability measure, since its domain is  $S$  instead of  $\mathcal{P}(S)$ . We build a probability measure from  $p$  by defining  $P(A) = \sum_{x \in A} p(x)$ . The axioms for a probability measure may be routinely verified. A similar construction works in the case where  $S$  is a countably infinite sample space. (Recall that a set  $S$  is *countably infinite* iff there exists a bijection  $f : \mathbb{N} \rightarrow S$ .)

**1.72. Remark.** In this section, we used counting techniques to solve basic probability questions. It is also possible to use probabilistic arguments to help solve counting problems. Examples of such arguments appear in §12.4 and §12.10.

## 1.13 Games of Chance

In this section, we use counting techniques to analyze two popular games of chance: power-ball lotteries and five-card poker.

**TABLE 1.3**

Analysis of the Powerball lottery.

Matches	Probability	Prize Value
0 white, 1 red	0.0145 or 1 in 69	\$3
1 white, 1 red	0.00788 or 1 in 127	\$4
2 white, 1 red	0.00134 or 1 in 745	\$7
3 white, 0 red	0.00344 or 1 in 291	\$7
3 white, 1 red	0.0000838 or 1 in 11,927	\$100
4 white, 0 red	0.0000702 or 1 in 14,254	\$100
4 white, 1 red	0.000001711 or 1 in 584,432	\$10,000
5 white, 0 red	$2.81 \times 10^{-7}$ or 1 in 3.56 million	\$200,000
5 white, 1 red	$6.844 \times 10^{-9}$ or 1 in 146 million	jackpot

**1.73. Example: Powerball.** A *powerball lottery* has two kinds of balls: white balls (numbered  $1, \dots, M$ ) and red balls (numbered  $1, \dots, R$ ). Each week, one red ball and a set of  $n$  distinct white balls are randomly chosen. Lottery players guess what the  $n$  white balls will be, and they also guess the red ball (called the “power ball”). Players win prizes based on how many balls they guess correctly. Players always win a prize for matching the red ball, even if they incorrectly guess all the white balls.

To analyze this lottery, let the sample space be

$$S = \{(T, x) : T \text{ is an } n\text{-element subset of } \{1, 2, \dots, M\} \text{ and } x \in \{1, 2, \dots, R\}\}.$$

Let  $(T_0, x_0)$  be a fixed and known element of  $S$  representing a given player’s lottery ticket. For  $0 \leq k \leq n$ , let  $A_k$  be the event  $\{(T, x) \in S : |T \cap T_0| = k, x \neq x_0\}$  in which the player matches exactly  $k$  white balls but misses the power ball. Let  $B_k$  be the event  $\{(T, x) \in S : |T \cap T_0| = k, x = x_0\}$  in which the player matches exactly  $k$  white balls and also matches the power ball. We have  $|S| = \binom{M}{n}R$  by the product rule. To build a typical element in  $A_k$ , we first choose  $k$  elements of  $T_0$ , then choose  $n - k$  elements of  $\{1, 2, \dots, M\} \sim T_0$ , then choose  $x \in \{1, 2, \dots, R\} \sim \{x_0\}$ . Thus,  $|A_k| = \binom{n}{k} \binom{M-n}{n-k} (R-1)$ , so

$$P(A_k) = \frac{\binom{n}{k} \binom{M-n}{n-k} (R-1)}{\binom{M}{n} R}.$$

Similarly,

$$P(B_k) = \frac{\binom{n}{k} \binom{M-n}{n-k} \cdot 1}{\binom{M}{n} R}.$$

In one version of this lottery, we have  $M = 55$ ,  $R = 42$ , and  $n = 5$ . The probabilities of certain events  $A_k$  and  $B_k$  are shown in Table 1.3 together with the associated prize amounts.

Now we turn to an analysis of five-card poker.

**1.74. Definition: Cards.** A *suit* is an element of the 4-element set  $Suits = \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$ . A *value* is an element of the 13-element set  $Values = \{2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A\}$ , where J, Q, K, and A stand for “jack,” “queen,” “king,” and “ace,” respectively. A *card* is an element of the set  $Deck = Values \times Suits$ .

Note that  $|Deck| = 13 \cdot 4 = 52$ , by the product rule. For instance,  $(A, \spadesuit) \in Deck$ . We often abbreviate this notation to  $A\spadesuit$ , and similarly for other cards.



**1.75. Definition: Poker Hands.** A (*five-card*) *poker hand* is a 5-element subset of Deck. Given such a hand  $H$ , let  $V(H)$  be the set of values that appear among the cards of  $H$ , and let  $S(H)$  be the set of suits that appear among the cards of  $H$ . For each  $x \in \text{Values}$ , let  $n_x(H)$  be the number of cards in  $H$  with value  $x$ .

**1.76. Example.**  $H = \{A\heartsuit, 3\clubsuit, 3\diamondsuit, J\diamondsuit, K\clubsuit\}$  is a five-card poker hand with  $V(H) = \{A, 3, J, K\}$ ,  $S(H) = \{\heartsuit, \clubsuit\}$ ,  $n_3(H) = 2$ ,  $n_A(H) = n_J(H) = n_K(H) = 1$ , and  $n_x(H) = 0$  for all  $x \notin V(H)$ .

We now study the sample space  $X$  consisting of all five-card poker hands. We know that  $|X| = \binom{52}{5} = 2,598,960$ . In poker, certain hands in  $X$  play a special role. We define these hands now.

**1.77. Definition: Special Card Hands.** Let  $H$  be a five-card poker hand.

- $H$  is a *four-of-a-kind hand* iff there exists  $x \in \text{Values}$  with  $n_x(H) = 4$ .
- $H$  is a *full house* iff there exist  $x, y \in \text{Values}$  with  $n_x(H) = 3$  and  $n_y(H) = 2$ .
- $H$  is a *three-of-a-kind hand* iff there exist  $x, y, z \in \text{Values}$  with  $y \neq z$ ,  $n_x(H) = 3$ , and  $n_y(H) = n_z(H) = 1$ .
- $H$  is a *two-pair hand* iff there exist  $x, y, z \in \text{Values}$  with  $y \neq z$ ,  $n_x(H) = 1$ , and  $n_y(H) = n_z(H) = 2$ .
- $H$  is a *one-pair hand* iff there exist distinct  $w, x, y, z \in \text{Values}$  with  $n_w(H) = 2$  and  $n_x(H) = n_y(H) = n_z(H) = 1$ .
- $H$  is a *straight* iff  $V(H)$  is one of the following sets:

$$\{A, 2, 3, 4, 5\} \text{ or } \{i, i+1, i+2, i+3, i+4\} \text{ for some } i \text{ with } 2 \leq i \leq 6$$

$$\text{or } \{7, 8, 9, 10, J\} \text{ or } \{8, 9, 10, J, Q\} \text{ or } \{9, 10, J, Q, K\} \text{ or } \{10, J, Q, K, A\}.$$

- $H$  is a *flush* iff  $|S(H)| = 1$ .
- $H$  is a *straight flush* iff  $H$  is a straight and a flush.
- $H$  is an *ordinary hand* iff  $H$  satisfies none of the above conditions.

**1.78. Example: Card Hands.**

- $\{5\spadesuit, 8\clubsuit, 5\diamondsuit, 5\clubsuit, 5\heartsuit\}$  is a four-of-a-kind hand.
- $\{J\spadesuit, 9\clubsuit, J\diamondsuit, J\clubsuit, 9\heartsuit\}$  is a full house.
- $\{J\spadesuit, 2\clubsuit, J\diamondsuit, J\clubsuit, 9\heartsuit\}$  is a three-of-a-kind hand.
- $\{2\spadesuit, 9\clubsuit, K\diamondsuit, 2\clubsuit, 9\heartsuit\}$  is a two-pair hand.
- $\{9\clubsuit, 10\diamondsuit, 10\heartsuit, A\clubsuit, 4\clubsuit\}$  is a one-pair hand.
- $\{7\clubsuit, 6\diamondsuit, 3\heartsuit, 5\clubsuit, 4\clubsuit\}$  is a straight that is not a flush.
- $\{10\heartsuit, 3\heartsuit, Q\heartsuit, J\heartsuit, 8\heartsuit\}$  is a flush that is not a straight.

**TABLE 1.4**

Probability of five-card poker hands.

Card Hand	Number	Probability
straight flush	40	$1.54 \times 10^{-5}$
four-of-a-kind	624	0.00024
full house	3744	0.00144
flush (not straight)	5108	0.001965
straight (not flush)	10,200	0.00392
three-of-a-kind	54,912	0.02113
two pair	123,552	0.04754
one pair	1,098,240	0.42257
none of the above	1,302,540	0.50117
TOTAL	2,598,960	1.00000

- $\{10\spadesuit, J\spadesuit, Q\spadesuit, K\spadesuit, A\spadesuit\}$  is a straight flush. (A straight flush such as this one, which “starts at 10 and ends at A,” is called a *royal flush*. There are four royal flushes, one for each suit.)
- $\{9\clubsuit, 10\diamondsuit, 7\heartsuit, A\clubsuit, 4\clubsuit\}$  is an ordinary hand.

We now compute the probability of the various five-card poker hands. This amounts to enumerating the hands of each type and dividing these counts by  $|X| = \binom{52}{5} = 2,598,960$ . Our results are summarized in Table 1.4. In each case, the desired counting result will follow from careful applications of the product rule. Less frequently occurring poker hands are more valuable in the game. So, for instance, a flush beats a straight. A full house beats both a straight and a flush separately, but is beaten by a straight flush.

- *Four-of-a-kind hands.* To build a typical four-of-a-kind hand  $H$ , first choose the value  $x$  that occurs 4 times in any of  $|\text{Values}| = 13$  ways. All four cards of this value must belong to  $H$ . Second, choose the fifth card of  $H$  in any of  $52 - 4 = 48$  ways. This gives  $13 \times 48 = 624$  four-of-a-kind hands. The sample hand above was constructed by choosing the value 5 followed by the card  $8\clubsuit$ .
- *Full house hands.* To build a typical full house  $H$ , first choose a value  $x \in \text{Values}$  to occur 3 times. This can be done in 13 ways. Second, choose 3 of the 4 cards of value  $x$  to appear in the hand. This can be done in  $\binom{4}{3} = 4$  ways. Third, choose a value  $y \in \text{Values} \sim \{x\}$  to occur twice in  $H$ . This can be done in 12 ways. Fourth, choose 2 of the 4 cards of value  $y$  to appear in the hand. This can be done in  $\binom{4}{2} = 6$  ways. The total is  $13 \cdot 4 \cdot 12 \cdot 6 = 3744$  full house hands. The sample hand above was constructed by choosing the value  $J$ , then the three cards  $\{J\spadesuit, J\diamondsuit, J\clubsuit\}$ , then the value 9, then the two cards  $\{9\clubsuit, 9\heartsuit\}$ .
- *Three-of-a-kind hands.* To build a typical three-of-a-kind hand  $H$ , first choose a value  $x \in \text{Values}$  to occur 3 times. This can be done in 13 ways. Second, choose 3 of the 4 cards of value  $x$  to appear in the hand. This can be done in  $\binom{4}{3} = 4$  ways. Third, choose a set of 2 values  $\{y, z\} \subseteq \text{Values} \sim \{x\}$  that will occur once each in  $H$ . This can be done in  $\binom{12}{2} = 66$  ways. Let the notation be such that  $y < z$  (where  $10 < J < Q < K < A$ ). Fourth, choose one of the 4 cards of value  $y$  to be in the hand in any of 4 ways. Fifth, choose one of the 4 cards of value  $z$  to be in the hand in any of 4 ways. The total is  $13 \cdot 4 \cdot 66 \cdot 4 \cdot 4 = 54,912$ . The sample hand above was constructed by choosing the value

$J$ , then the three cards  $\{J\spadesuit, J\diamondsuit, J\clubsuit\}$ , then the values  $\{2, 9\}$ , then the card  $2\clubsuit$ , and then the card  $9\heartsuit$ .

- *Two-pair hands.* To build a typical two-pair hand  $H$ , first choose a set of two values  $\{x, y\} \in \text{Values}$  to occur twice each. This can be done in  $\binom{13}{2} = 78$  ways. Let the notation be such that  $x < y$ . Second, choose a set of two cards of value  $x$  in any of  $\binom{4}{2} = 6$  ways. Third, choose a set of two cards of value  $y$  in any of  $\binom{4}{2} = 6$  ways. Fourth, choose the last card in the hand. Since this card cannot have value  $x$  or  $y$ , the number of possibilities here is  $52 - 8 = 44$ . The total is  $78 \cdot 6 \cdot 6 \cdot 44 = 123,552$ . The sample hand above was constructed by choosing the values  $\{2, 9\}$ , then the cards  $\{2\spadesuit, 2\clubsuit\}$ , then the cards  $\{9\clubsuit, 9\heartsuit\}$ , then the card  $K\diamondsuit$ .
- *One-pair hands.* To build a typical one-pair hand  $H$ , first choose a value  $w$  to occur twice in the hand. This can be done in 13 ways. Second, choose a set of two cards of value  $w$  in any of  $\binom{4}{2} = 6$  ways. Third, choose a set  $\{x, y, z\} \subset \text{Values} \sim \{x\}$  (where  $x < y < z$ ) in any of  $\binom{12}{3} = 220$  ways. Fourth, choose a card of value  $x$  in 4 ways. Fifth, choose a card of value  $y$  in 4 ways. Sixth, choose a card of value  $z$  in 4 ways. The total is  $13 \cdot 6 \cdot 220 \cdot 4 \cdot 4 \cdot 4 = 1,098,240$ . The sample hand above was constructed by choosing the value  $w = 10$ , then the cards  $\{10\diamondsuit, 10\heartsuit\}$ , then the values  $\{4, 9, A\}$ , then the card  $4\clubsuit$ , then the card  $9\clubsuit$ , then the card  $A\clubsuit$ .
- *Straight hands.* To build a typical straight  $H$ , first choose one of the ten allowable sets  $V(H)$  in the definition of a straight. Then, for each of the five distinct values in  $V(H)$ , taken in increasing order, choose a suit for the card of that value. This can be done in 4 ways for each value. The total is  $10 \cdot 4^5 = 10,240$ . The sample hand above was constructed by choosing the value set  $V(H) = \{3, 4, 5, 6, 7\}$ , then the suit  $\heartsuit$  for the 3, then the suit  $\clubsuit$  for the 4, then the suit  $\clubsuit$  for the 5, then the suit  $\diamondsuit$  for the 6, and then the suit  $\clubsuit$  for the 7. In the table entry for straights, we subtract the number of straight flushes (namely 40, as shown below) so that the entries in the table will be pairwise disjoint subsets of  $X$ .
- *Flush hands.* To build a typical flush  $H$ , first choose the one-element set  $S(H)$  in any of  $\binom{4}{1} = 4$  ways. Then choose the five-element set  $V(H)$  in any of  $\binom{13}{5}$  ways.  $H$  is now completely determined since all cards in  $H$  have the same suit. The total is therefore  $4 \cdot \binom{13}{5} = 5148$ . The sample hand above was constructed by choosing  $S(H) = \{\heartsuit\}$ , then  $V(H) = \{3, 8, 10, J, Q\}$ . In the table entry for flushes, we subtract the number of straight flushes (namely 40, as shown below) so that the entries in the table will be pairwise disjoint subsets of  $X$ .
- *Straight flushes.* To build a typical straight flush  $H$ , first choose one of the ten allowable sets  $V(H)$  in the definition of a straight. Then choose one of the four suits to be the common suit of all cards in  $H$ . The total is  $10 \cdot 4 = 40$ . The sample hand above was constructed by choosing  $V(H) = \{10, J, Q, K, A\}$  and then  $S(H) = \{\spadesuit\}$ .
- *Ordinary hands.* To count ordinary hands, one can subtract the total of the preceding counts from  $|X|$ . However, the answer can also be obtained directly from the product rule as follows. To build an ordinary hand  $H$ , first choose the value set  $V(H)$ . We must have  $|V(H)| = 5$  to avoid hands such as two-pair, full house, etc. Also we must avoid the ten special choices of  $V(H)$  in the definition of straight (all of which are five-element sets). We conclude that  $V(H)$  can be chosen in  $\binom{13}{5} - 10 = 1277$  ways. Write  $V(H) = \{v_1, v_2, v_3, v_4, v_5\}$ , where  $v_1 < v_2 < v_3 < v_4 < v_5$ . For each  $v_i$  in turn, choose the suit for the card of that value in any of 4 ways. This would give  $4^5$  choices, but we must avoid the four choice sequences in which all  $v_i$ 's are assigned the same suit (which

would lead to a flush). So there are only  $4^5 - 4 = 1020$  ways to assign suits to the chosen values. The hand is now completely determined, so the total number of ordinary hands is  $1277 \cdot 1020 = 1,302,540$ . The sample hand above was constructed by choosing  $V(H) = \{4, 7, 9, 10, A\}$ , then  $\clubsuit$  as the suit for the 4,  $\heartsuit$  as the suit for the 7,  $\clubsuit$  as the suit for the 9,  $\diamondsuit$  as the suit for the 10, and  $\clubsuit$  as the suit for the ace.

## 1.14 Conditional Probability and Independence

Suppose that, in a certain random experiment, we are told that a particular event has occurred. Given this additional information, we can recompute the probability of other events occurring. This leads to the notion of conditional probability.

**1.79. Definition: Conditional Probability.** Suppose  $A$  and  $B$  are events in some sample space  $S$  such that  $P(B) > 0$ . The *conditional probability of  $A$  given  $B$* , denoted  $P(A|B)$ , is defined by setting

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

In the case where  $S$  is a finite set of equally likely outcomes, we have  $P(A|B) = |A \cap B|/|B|$ . This conditional probability need not have any relation to the *unconditional* probability of  $A$ , which is  $P(A) = |A|/|S|$ .

**1.80. Example: Dice Rolling.** Consider the experiment of rolling a fair die twice. What is the probability of getting a sum of 7 or 11, given that the second roll comes up 5? Here, the sample space is  $S = \{1, 2, 3, 4, 5, 6\}^2$ . Let  $A$  be the event of getting a sum of 7 or 11, and let  $B$  be the event that the second die shows 5. We have  $P(B) = 1/6$ , and we saw earlier that  $P(A) = 2/9$ . Listing outcomes, we see that  $A \cap B = \{(2, 5), (6, 5)\}$ , so  $P(A \cap B) = 2/36 = 1/18$ . Therefore, the required conditional probability is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/18}{1/6} = 1/3 > 2/9 = P(A).$$

On the other hand, let  $C$  be the event that the second roll comes up 4. Here  $A \cap C = \{(3, 4)\}$ , so

$$P(A|C) = \frac{1/36}{6/36} = 1/6 < 2/9 = P(A).$$

Next, let  $D$  be the event that the first roll is an odd number. Then

$$A \cap D = \{(1, 6), (3, 4), (5, 2), (5, 6)\},$$

so

$$P(A|D) = \frac{4/36}{18/36} = 2/9 = P(A).$$

These examples show that the conditional probability of  $A$  given some other event can be greater than, less than, or equal to the unconditional probability of  $A$ .

**1.81. Example: Balls in Urns.** Suppose an urn contains  $r$  red balls and  $b$  blue balls, where  $r, b \geq 2$ . Consider an experiment in which two balls are drawn from the urn in succession, without replacement. What is the probability that the first ball is red, given

that the second ball is blue? We take the sample space to be the set  $S$  of all words  $w_1w_2$ , where  $w_1 \neq w_2$  and

$$w_1, w_2 \in \{1, 2, \dots, r, r+1, \dots, r+b\}.$$

Here, the numbers 1 through  $r$  represent red balls and the numbers  $r+1$  through  $r+b$  represent blue balls. The event of drawing a red ball first is the subset

$$A = \{w_1w_2 : 1 \leq w_1 \leq r\}.$$

The event of drawing a blue ball second is the subset

$$B = \{w_1w_2 : r+1 \leq w_2 \leq r+b\}.$$

By the product rule,  $|S| = (r+b)(r+b-1)$ ,  $|A| = r(r+b-1)$ ,  $|B| = b(r+b-1)$ , and  $|A \cap B| = rb$ . The conditional probability of  $A$  given  $B$  is

$$P(A|B) = P(A \cap B)/P(B) = r/(r+b-1).$$

In contrast, the unconditional probability of  $A$  is

$$P(A) = |A|/|S| = r/(r+b).$$

The conditional probability is slightly higher than the unconditional probability; intuitively, we are more likely to have gotten a red ball first if we know the second ball was not red. The probability that the second ball is blue, given that the first ball is red, is

$$P(B|A) = P(B \cap A)/P(A) = b/(r+b-1).$$

Note that  $P(B|A) \neq P(A|B)$  (unless  $r = b$ ).

**1.82. Example: Card Hands.** What is the probability that a 5-card poker hand is a full house, given that the hand is void in clubs (i.e., no card in the hand is a club)? Let  $A$  be the event of getting a full house, and let  $B$  be the event of being void in clubs. We have  $|B| = \binom{39}{5} = 575,757$  since we must choose a five-element subset of the  $52 - 13 = 39$  non-club cards. Next, we must compute  $|A \cap B|$ . To build a full house hand using no clubs, make the following choices: first, choose a value to occur three times (13 ways); second, choose the suits for this value (1 way, as clubs are forbidden); third, choose a value to occur twice (12 ways); fourth, choose the suits for this value ( $\binom{3}{2} = 3$  ways). By the product rule,  $|A \cap B| = 13 \cdot 1 \cdot 12 \cdot 3 = 468$ . Accordingly, the probability we want is  $P(A|B) = 468/575,757 \approx 0.000813$ .

Next, what is the probability of getting a full house, given that the hand has at least two cards of the same value? Let  $C$  be the event that at least two cards in the hand have the same value; we seek  $P(A|C) = P(A \cap C)/P(C) = |A \cap C|/|C|$ . The numerator here can be computed quickly: since  $A \subseteq C$ , we have  $A \cap C = A$  and hence  $|A \cap C| = |A| = 3744$  (see Table 1.4). To compute the denominator, let us first enumerate  $X \sim C$ , where  $X$  is the full sample space of all five-card poker hands. Note that  $X \sim C$  occurs iff all five cards in the hand have different values. Choose these values ( $\binom{13}{5}$  ways), and then choose suits for each card (4 ways each). By the product rule,  $|X \sim C| = 1,317,888$ . So

$$|C| = |X| - |X \sim C| = 1,281,072.$$

The desired conditional probability is

$$P(A|C) = \frac{3744}{1,281,072} \approx 0.00292.$$

In some situations, the knowledge that a particular event  $D$  occurs does not change the probability that another event  $A$  will occur. For instance, events  $D$  and  $A$  in the dice rolling example 1.80 have this property because  $P(A|D) = P(A)$ . Writing out the definition of  $P(A|D)$  and multiplying by  $P(D)$ , we see that the stated property is equivalent to  $P(A \cap D) = P(A)P(D)$  (assuming  $P(D) > 0$ ). This suggests the following definition, which is valid even when  $P(D) = 0$ .

**1.83. Definition: Independence of Two Events.** Two events  $A$  and  $D$  are called *independent* iff

$$P(A \cap D) = P(A)P(D).$$

Unlike the definition of conditional probability, this definition is symmetric in  $A$  and  $D$ . So,  $A$  and  $D$  are independent iff  $D$  and  $A$  are independent. As indicated above, when  $P(D) > 0$ , independence of  $A$  and  $D$  is equivalent to  $P(A|D) = P(A)$ . Similarly, when  $P(A) > 0$ , independence of  $A$  and  $D$  is equivalent to  $P(D|A) = P(D)$ . So, when considering two independent events of positive probability, knowledge that either event has occurred gives us no new information about the probability of the other event occurring.

**1.84. Definition: Independence of a Collection of Events.** Suppose  $A_1, \dots, A_n$  are events. This list of events is called *independent* iff for all choices of indices  $i_1 < i_2 < \dots < i_k \leq n$ ,

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_k}).$$

**1.85. Example.** Let  $S = \{a, b, c, d\}$ , and suppose each outcome in  $S$  occurs with probability  $1/4$ . Define events  $B = \{a, b\}$ ,  $C = \{a, c\}$ , and  $D = \{a, d\}$ . One verifies immediately that  $B$  and  $C$  are independent;  $B$  and  $D$  are independent; and  $C$  and  $D$  are independent. However, the triple of events  $B, C, D$  is *not* independent, because

$$P(B \cap C \cap D) = P(\{a\}) = 1/4 \neq 1/8 = P(B)P(C)P(D).$$

**1.86. Example: Coin Tossing.** Suppose we toss a fair coin 5 times. Take the sample space to be  $S = \{H, T\}^5$ . Let  $A$  be the event that the first and last toss agree; let  $B$  be the event that the third toss is tails; let  $C$  be the event that there are an odd number of heads. Routine counting arguments show that  $|S| = 2^5 = 32$ ,  $|A| = 2^4 = 16$ ,  $|B| = 2^4 = 16$ ,  $|C| = \binom{5}{1} + \binom{5}{3} + \binom{5}{5} = 16$ ,  $|A \cap B| = 2^3 = 8$ ,  $|A \cap C| = 2(\binom{3}{1} + \binom{3}{3}) = 8$ ,  $|B \cap C| = \binom{4}{1} + \binom{4}{3} = 8$ , and  $|A \cap B \cap C| = 4$ . It follows that

$$P(A \cap B) = P(A)P(B); \quad P(A \cap C) = P(A)P(C); \quad P(B \cap C) = P(B)P(C);$$

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Thus, the triple of events  $(A, B, C)$  is independent.

We often assume that unrelated physical events are independent (in the mathematical sense) to help us construct a probability model. The next example illustrates this process.

**1.87. Example: Tossing an Unfair Coin.** Consider a random experiment in which we toss an unbalanced coin  $n$  times in a row. Suppose that the coin comes up heads with probability  $q$  and tails with probability  $1 - q$ , and that successive coin tosses are unrelated to one another. Let the sample space be  $S = \{H, T\}^n$ . Since the coin is unfair, it is not appropriate to assume that every point of  $S$  occurs with equal probability. Given an outcome  $w = w_1 w_2 \dots w_n \in S$ , what should the probability  $p(w)$  be? Consider an example where  $n = 5$  and  $w = \text{HHTHT}$ . Consider the five events  $B_1 = \{z \in S : z_1 = H\}$ ,  $B_2 = \{z \in S : z_2 = H\}$ ,  $B_3 = \{z \in S : z_3 = T\}$ ,  $B_4 = \{z \in S : z_4 = H\}$ , and  $B_5 = \{z \in S : z_5 = T\}$ . Our physical assumptions suggest that  $B_1, \dots, B_5$  should be independent events (since different

tosses of the coin are unrelated),  $P(B_1) = P(B_2) = P(B_4) = q$ , and  $P(B_3) = P(B_5) = 1 - q$ . Since  $B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5 = \{w\}$ , the definition of independence leads to

$$p(w) = P(B_1 \cap \cdots \cap B_5) = P(B_1)P(B_2) \cdots P(B_5) = qq(1-q)q(1-q) = q^3(1-q)^2.$$

Similar reasoning shows that if  $w = w_1 w_2 \cdots w_n \in S$  is an outcome consisting of  $k$  heads and  $n - k$  tails (arranged in one particular order), then we should define  $p(w) = q^k(1-q)^{n-k}$ . Next, define  $P(A) = \sum_{w \in A} p(w)$  for every event  $A \subseteq S$ . For example, let  $A_k$  be the event that we get  $k$  heads and  $n - k$  tails (in any order). Note that  $|A_k| = |\mathcal{R}(H^k T^{n-k})| = \binom{n}{k}$ , and  $p(w) = q^k(1-q)^{n-k}$  for each  $w \in A_k$ . It follows that

$$P(A_k) = \binom{n}{k} q^k (1-q)^{n-k}.$$

We have not yet checked that  $P(S) = 1$ , which is one of the requirements in the definition of a probability measure. This fact can be deduced from the binomial theorem (discussed in §2.2), as follows. Since  $S$  is the disjoint union of  $A_0, A_1, \dots, A_n$ , we have

$$P(S) = \sum_{k=0}^n \binom{n}{k} q^k (1-q)^{n-k}.$$

By the binomial theorem 2.14, the right side is  $(q + [1 - q])^n = 1^n = 1$ .

## Summary

We end each chapter by summarizing some of the main definitions and results discussed in the chapter.

- *Notation.* Factorials:  $0! = 1$  and  $n! = n \times (n-1) \times \cdots \times 1$ .  
Binomial coefficients:  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$  for  $0 \leq k \leq n$ ;  $\binom{n}{k} = 0$  otherwise.  
Multinomial coefficients: Given  $n_1, \dots, n_k \geq 0$  and  $N = n_1 + \cdots + n_k$ ,  $\binom{N}{n_1, \dots, n_k} = \frac{N!}{n_1! n_2! \cdots n_k!}$ .  
Rearrangements:  $\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  is the set of all words consisting of  $n_i$  copies of  $a_i$ .
- *Basic Counting Rules.* Sum Rule: If  $A_1, \dots, A_k$  are pairwise disjoint finite sets, then  $|A_1 \cup \cdots \cup A_k| = |A_1| + \cdots + |A_k|$ .  
Union Rule: If  $A$  and  $B$  are arbitrary finite sets, then  $|A \cup B| = |A| + |B| - |A \cap B|$ .  
Difference Rule: If  $A \subseteq B$  and  $B$  is finite, then  $|B \setminus A| = |B| - |A|$ .  
Product Rule: If  $A_1, \dots, A_k$  are arbitrary finite sets, then  $|A_1 \times \cdots \times A_k| = |A_1| \cdots |A_k|$ .  
Bijection Rule: If there is a bijection  $f: A \rightarrow B$ , then  $|A| = |B|$ .
- *Counting Words.* Let  $A$  be an  $n$ -letter alphabet.  
There are  $n^k$  words of length  $k$  using letters from  $A$ .  
If the letters must be distinct, there are  $n!/(n-k)!$  words of length  $k \leq n$ .  
There are  $n!$  permutations of all the letters in  $A$ .  
There are  $\binom{n_1 + \cdots + n_k}{n_1, \dots, n_k}$  words in  $\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$ .
- *Counting Sets and Multisets.*  
The number of  $k$ -element subsets of an  $n$ -element set is the binomial coefficient  $\binom{n}{k}$ .  
The total number of subsets of an  $n$ -element set is  $2^n$ .  
The number of  $k$ -element multisets using  $n$  available objects is  $\binom{k+n-1}{k, n-1}$ .

- *Counting Functions.* Let  $|X| = a$  and  $|Y| = b$ .  
There are  $b^a$  functions mapping  $X$  into  $Y$ .  
For  $a \leq b$ , there are  $b!/(b-a)!$  injections from  $X$  to  $Y$ .  
If  $a = b$ , there are  $a!$  bijections from  $X$  onto  $Y$ .
- *Counting Lattice Paths.* There are  $\binom{a+b}{a,b}$  lattice paths from  $(0,0)$  to  $(a,b)$ .  
There are  $\binom{n_1+\dots+n_d}{n_1,\dots,n_d}$  lattice paths in  $\mathbb{R}^d$  from the origin to  $(n_1, n_2, \dots, n_d)$ .  
The number of paths from  $(0,0)$  to  $(n,n)$  that never go below  $y = x$  is the Catalan number

$$C_n = \frac{1}{n+1} \binom{2n}{n} = \frac{1}{2n+1} \binom{2n+1}{n} = \binom{2n}{n,n} - \binom{2n}{n+1,n-1}.$$

This can be proved using a reflection bijection to convert paths ending at  $(n,n)$  that do go below  $y = x$  to arbitrary paths from  $(0,0)$  to  $(n+1, n-1)$ .

- *Compositions.* A composition of  $n$  is an ordered sequence  $(\alpha_1, \dots, \alpha_k)$  of positive integers that sum to  $n$ . There are  $2^{n-1}$  compositions of  $n$ . There are  $\binom{n-1}{k-1}$  compositions of  $n$  with  $k$  parts.
- *Probability Definitions.* A *sample space* is the set  $S$  of outcomes for some random experiment. An *event* is a subset of the sample space. When all outcomes in  $S$  are equally likely, the *probability* of an event  $A$  is  $P(A) = |A|/|S|$ . The *conditional probability of A given B* is  $P(A|B) = P(A \cap B)/P(B)$ , when  $P(B) > 0$ . Events  $A$  and  $B$  are *independent* iff  $P(A \cap B) = P(A)P(B)$ .

## Exercises

- 1.88.** (a) How many numbers between 1 and 1000 are divisible by 5 or 7? (b) How many such numbers are divisible by 5 or 7, but not both?
- 1.89.** How many three-digit numbers: (a) do not contain the digits 5 or 7; (b) contain the digits 5 and 7; (c) contain the digits 5 or 7; (d) contain 5 or 7, but not both?
- 1.90.** How many seven-digit phone numbers do not begin with one of the prefixes 1, 911, 411, or 555?
- 1.91.** How many  $n$ -letter words over the alphabet  $\{0, 1\}$  use both the symbols 0 and 1?
- 1.92.** (a) How many four-letter words  $w$  using an  $n$ -letter alphabet satisfy  $w_i \neq w_{i+1}$  for  $i = 1, 2, 3$ ? (b) How many of the words in (a) also satisfy  $w_4 \neq w_1$ ?
- 1.93.** A key for the *DES encryption system* is a binary word of length 56. A key for a *permutation cipher* is a permutation of the 26-letter English alphabet. Which encryption system has more keys?
- 1.94.** A key for the *AES encryption system* is a binary word of length 128. Suppose we try to decrypt an AES message by exhaustively trying every possible key. Assume six billion computers are running in parallel, where each computer can test one trillion keys per second. Estimate the number of years required for this attack to search the entire space of keys.



**1.95.** A pizza shop offers ten toppings. How many pizzas can be ordered with: (a) three different toppings; (b) up to three different toppings; (c) three toppings, with repeats allowed; (d) four different toppings, but pepperoni and sausage cannot be ordered together?

**1.96.** How many lattice paths from  $(0, 0)$  to  $(7, 5)$  pass through the point  $(2, 3)$ ?

**1.97.** How many  $n$ -letter words contain: (a) only vowels; (b) no vowels; (c) at least one vowel; (d) alternating vowels and consonants; (e) two vowels and  $n - 2$  consonants? (The vowels are A, E, I, O, and U.)

**1.98.** How many four-digit even numbers contain the digit 5 but not the digit 2?

**1.99.** A *palindrome* is a word  $w = w_1w_2 \cdots w_k$  that reads the same in reverse, i.e.,  $w_1w_2 \cdots w_k = w_k \cdots w_2w_1$ . Count the number of  $k$ -letter palindromes using letters from an  $n$ -letter alphabet.

**1.100.** Explicitly list the following objects: (a) all 4-letter words using the alphabet  $\{0, 1\}$ ; (b) all permutations of  $\{a, b, c, d\}$ ; (c) all 2-permutations of  $\{u, v, w, x, y\}$ ; (d) all words in  $\mathcal{R}(x^2y^2z^1)$ .

**1.101.** Explicitly list the following objects: (a) all bijections from  $\{1, 2, 3\}$  to  $\{i, j, k\}$ ; (b) all surjections from  $\{1, 2, 3\}$  to  $\{0, 1\}$ ; (c) all injections from  $\{a, b\}$  to  $\{c, d, e, f\}$ .

**1.102.** Explicitly list the following objects: (a) all subsets of  $\{0, 1, 2\}$ ; (b) all three-element subsets of  $\{1, 2, 3, 4, 5\}$ ; (c) all three-element multisets using the alphabet  $\{a, b, c\}$ .

**1.103.** Explicitly list the following objects: (a) all compositions of 4; (b) all compositions of 7 with exactly three parts; (c) all lattice paths from  $(0, 0)$  to  $(4, 2)$ ; (d) all Dyck paths of order 4.

**1.104.** Draw pictures of all compositions of 5. For each composition, determine the associated word in  $\{0, 1\}^4$  constructed in the proof of 1.41.

**1.105.** How many lattice paths start at  $(0, 0)$  and end on the line  $x + y = n$ ?

**1.106.** Let  $r$  be the bijection in the proof of 1.56. Compute

$$r(\text{NNEEEENNNNEEEENN}) \text{ and } r^{-1}(\text{NENEENNEEEENNEENN}).$$

**1.107.** Draw all the non-Dyck lattice paths from  $(0, 0)$  to  $(3, 3)$  and compute their images under the reflection map  $r$  from the proof of 1.56.

**1.108.** A *bit* is one of the symbols 0 or 1. Find the minimum  $k$  such that every printable character on a standard computer keyboard can be encoded by a distinct bit string of length exactly  $k$ . Does the answer change if we allow nonempty bit strings of length *at most*  $k$ ?

**1.109.** Ten lollipops are to be distributed to four children. All lollipops of the same color are considered identical. How many distributions are possible if (a) all lollipops are red; (b) all lollipops have different colors; (c) there are four red and six blue lollipops? (d) What are the answers if each child must receive at least one lollipop?

**1.110.** Given a positive integer  $n$ , let the prime factorization of  $n$  be  $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ , where each  $e_j > 0$  and the  $p_j$  are distinct primes. How many positive divisors does  $n$  have? How many divisors does  $n$  have in  $\mathbb{Z}$ ?

**1.111.** (a) Given  $k$  and  $N$ , count the number of weakly increasing sequences  $(i_1 \leq i_2 \leq \cdots \leq i_k)$  with  $1 \leq i_j \leq N$  for all  $j$ . (b) Count the number of strictly decreasing sequences  $(i_1 > i_2 > \cdots > i_k)$  with  $1 \leq i_j \leq N$  for all  $j$ . (c) For a fixed choice of  $k$ , count the number of permutations  $w$  of  $N$  objects such that

$$w_1 < w_2 < \cdots < w_k > w_{k+1} < w_{k+2} < \cdots < w_N. \quad (1.2)$$

(d) How many permutations satisfy (1.2) for some  $k < N$ ?

**1.112. Euler's  $\phi$  Function.** For each  $n \geq 1$ , let  $\Phi(n)$  be the set of integers  $k$  between 1 and  $n$  such that  $\gcd(k, n) = 1$ , and let  $\phi(n) = |\Phi(n)|$ . (a) Compute  $\Phi(n)$  and  $\phi(n)$  for  $1 \leq n \leq 12$ . (b) Compute  $\phi(p)$  for  $p$  prime. (c) Compute  $\phi(p^e)$  for  $p$  prime and  $e > 1$ . (Exercise 1.150 shows how to compute  $\phi(n)$  for any  $n$ .)

**1.113.** (a) How many 4-element subsets of  $\{1, 2, \dots, 11\}$  contain no two consecutive integers? (b) Given  $d, k, n$ , how many  $k$ -element subsets  $S$  of  $\{1, 2, \dots, n\}$  are such that any two distinct elements of  $S$  differ by at least  $d$ ?

**1.114.** (a) How many anagrams of 'MISSISSIPPI' are there? (b) How many of these anagrams begin and end with P? (c) In how many of these anagrams are the two P's adjacent? (d) In how many of these anagrams are no two I's adjacent?

**1.115.** A *two-to-one function* is a function  $f : X \rightarrow Y$  such that for every  $y \in Y$ , there exist *exactly two* elements  $x_1, x_2 \in X$  with  $f(x_1) = y = f(x_2)$ . How many two-to-one functions are there from a  $2n$ -element set to an  $n$ -element set?

**1.116.** A *monomial* in  $N$  variables is a term of the form  $x_1^{k_1} x_2^{k_2} \cdots x_N^{k_N}$ , where each  $k_i \geq 0$ . The *degree* of this monomial is  $k_1 + k_2 + \cdots + k_N$ . How many monomials in  $N$  variables have degree (a) exactly  $d$ ; (b) at most  $d$ ?

**1.117.** How many multisets (of any size) can be formed from an  $n$ -letter alphabet if each letter can appear at most  $k$  times in the multiset?

**1.118.** Two fair dice are rolled. Find the probability that: (a) the same number appears on both dice; (b) the sum of the numbers rolled is 8; (c) the sum of the numbers rolled is divisible by 3; (d) the two numbers rolled differ by 1.

**1.119.** In blackjack, you have been dealt two cards from a shuffled 52-card deck:  $9\heartsuit$  and  $6\clubsuit$ . Find the probability that drawing one more card will cause the sum of the three card values to go over 21. (Here, an ace counts as 1 and other face cards count as 10.)

**1.120.** Find the probability that a random 5-letter word: (a) has no repeated letters; (b) contains no vowels; (c) is a palindrome.

**1.121.** A company employs ten men (one of whom is Bob) and eight women (one of whom is Alice). A four-person committee is randomly chosen. Find the probability that the committee: (a) consists of all men; (b) consists of two men and two women; (c) does not have both Alice and Bob as members.

**1.122.** A fair coin is tossed ten times. (a) Find the probability of getting exactly seven heads. (b) Find the probability of getting at least two heads. (c) Find the probability of getting exactly seven heads, given that the number of heads was prime.

**1.123.** A fair die is tossed ten times. What is the probability that, in these ten tosses, 1 comes up 5 times, 3 comes up 2 times, and 6 comes up 3 times?

**1.124.** Ten balls are drawn (without replacement) from an urn containing 40 red, 30 blue, and 30 white balls. (a) What is the probability that no blue balls are drawn? (b) What is the probability of getting 4 red, 3 blue, and 3 white balls? (c) What is the probability that all ten balls have the same color? (d) Answer the same questions assuming the balls are drawn with replacement.

**1.125.** Urn A contains two red balls and three black balls. Urn B contains one red ball and four black balls. Urn C contains four red balls and one black ball. A ball is randomly chosen from each of the three urns. Find the probability that all three balls are the same color.

**1.126.** Consider the three urns from 1.125 (with five balls in each urn). An urn is selected at random, and then one ball is selected from that urn. What is the probability that: (a) the ball is black, given that urn B was chosen; (b) the ball is black; (c) urn B was chosen, given that the ball was black?

**1.127.** A fair coin is tossed three times. (a) Describe the sample space. (b) Consider the following events.  $A$ : second toss is tails;  $B$ : first and last tosses agree;  $C$ : all tosses are the same;  $D$ : the number of heads is odd. Describe each event as a subset of the sample space. (c) Which pairs of events from  $\{A, B, C, D\}$  are independent? (d) Is the triple of events  $A, B, D$  independent? Explain.

**1.128.** Let the prime factorization of  $n!$  be  $p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ . Prove that  $e_i = \sum_{k=1}^{\infty} \lfloor n/p_i^k \rfloor$ . (The notation  $\lfloor x \rfloor$  denotes the greatest integer not exceeding the real number  $x$ .) Hence determine the number of trailing zeroes in the decimal notation for  $100!$ .

**1.129.** Find a bijection on  $\text{Comp}_n$  that maps compositions with  $k$  parts to compositions with  $n + 1 - k$  parts for all  $k$ .

**1.130.** (a) How many numbers between one and one million contain the digit 7? (b) If one writes down the numbers from one to one million, how often will one write the digit 7? (c) What are the answers to (a) and (b) if 7 is replaced by 0?

**1.131.** A *relation from  $X$  to  $Y$*  is any subset of  $X \times Y$ . Suppose  $X$  has  $n$  elements and  $Y$  has  $k$  elements. (a) How many relations from  $X$  to  $Y$  are there? (b) How many relations  $R$  satisfy the following property: for each  $y \in Y$ , there exists at most one  $x \in X$  with  $(x, y) \in R$ ?

**1.132.** Suppose we play five-card poker using a 51-card deck in which the queen of spades has been removed. Compute the probabilities of the poker hands in Table 1.4 relative to this deck.

**1.133.** Suppose we play five-card poker using two identical decks mixed together. Compute the probabilities of the poker hands in Table 1.4 in this situation. Also compute the probability of a “five-of-a-kind” hand, which is a poker hand  $H$  such that  $|V(H)| = 1$ .

**1.134.** Consider a five-card poker hand dealt from a 52-card deck. (a) What is the probability that the hand contains only red cards (i.e., hearts and diamonds)? (b) What is the probability that the hand contains exactly two eights? (c) What is the probability that the hand contains only numerical cards (i.e., ace, jack, queen, and king may not appear)?

**1.135.** Consider a five-card poker hand dealt from a 52-card deck. (a) What is the probability that the hand is a flush, given that the hand contains no clubs? (b) What is the probability that the hand contains at least one card from each of the four suits? (c) What is the probability of getting a two-pair hand, given that at least two cards in the hand have the same value?

**1.136.** Let  $K$  be the event that a five-card poker hand contains the card  $K\heartsuit$ . Find the conditional probability of each event in Table 1.4, given  $K$ . Which of these events are independent of  $K$ ?

**1.137. Texas Hold 'em.** In a popular version of poker, a player is dealt an *ordered* sequence of seven distinct cards from a 52-card deck. We model this situation using the sample space

$$S = \{(C_1, C_2, \dots, C_7) : C_i \in \text{Deck}, C_i \neq C_j \text{ for } i \neq j\}.$$

(The last five cards in this sequence are “community cards” shared with other players. In this exercise we concentrate on a single player, so we ignore this aspect of the game.) The player uses these seven cards to form the best possible five-card poker hand (cf. Table 1.4). For example, if we were dealt the hand

$$(4\heartsuit, 7\clubsuit, 3\diamondsuit, 9\clubsuit, 5\clubsuit, 6\clubsuit, Q\clubsuit),$$

we would have a flush (the five club cards) since this beats the straight (3,4,5,6,7 of various suits). (a) Compute  $|S|$ . (b) What is the probability of getting 4-of-a-kind? (c) What is the probability of getting a flush? (d) What is the probability of getting 4-of-a-kind, given  $C_1 = 3\heartsuit$  and  $C_2 = 3\spadesuit$ ? (e) What is the probability of getting a flush, given  $C_1 = 5\diamondsuit$  and  $C_2 = 9\diamondsuit$ ?

**1.138.** Prove that the following conditions are equivalent for any sets  $A$  and  $B$ : (a)  $A \subseteq B$ ; (b)  $A \cap B = A$ ; (c)  $A \cup B = B$ ; (d)  $A \sim B = \emptyset$ .

**1.139.** Prove that if  $A$  and  $B$  are unequal nonempty sets, then  $A \times B \neq B \times A$ .

**1.140.** Use the binary union rule 1.4 to prove that for all finite sets  $X, Y, Z$ ,

$$|X \cup Y \cup Z| = |X| + |Y| + |Z| - |X \cap Y| - |X \cap Z| - |Y \cap Z| + |X \cap Y \cap Z|.$$

**1.141.** (a) For fixed  $k$ , prove that  $\lim_{n \rightarrow \infty} \frac{n!}{(n-k)!n^k} = 1$ . (b) Give a probabilistic interpretation of this result.

**1.142.** Let  $f : \mathcal{P}(X) \rightarrow \{0, 1\}^n$  be the bijection in 1.38. Given two words  $v, w \in \{0, 1\}^n$ , define words  $v \wedge w$ ,  $v \vee w$ , and  $\neg v$  by setting  $(v \wedge w)_i = \min(v_i, w_i)$ ,  $(v \vee w)_i = \max(v_i, w_i)$ , and  $(\neg v)_i = 1 - v_i$  for all  $i \leq n$ . Prove that for all  $S, T \subseteq X$ ,  $f(S \cap T) = f(S) \wedge f(T)$ ,  $f(S \cup T) = f(S) \vee f(T)$ ,  $f(X \sim S) = \neg f(S)$ ,  $f(\emptyset) = 00 \cdots 0$ , and  $f(X) = 11 \cdots 1$ .

**1.143.** Let  $A, B, C$  be events in a probability space  $S$ . Assume  $A$  and  $C$  are independent, and  $B$  and  $C$  are independent. (a) Give an example where  $A \cup B$  and  $C$  are not independent. (b) Prove that  $A \cup B$  and  $C$  are independent if  $A$  and  $B$  are disjoint. (c) Must  $A \cap B$  and  $C$  be independent? Explain.

**1.144. Properties of Injections.** Prove the following statements about injective functions. (a) If  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$  are injective, then  $g \circ f$  is injective. (b) If  $g \circ f$  is injective, then  $f$  is injective but  $g$  may not be. (c)  $f : X \rightarrow Y$  is injective iff for all  $W$  and all  $g, h : W \rightarrow X$ ,  $f \circ g = f \circ h$  implies  $g = h$ .

**1.145. Properties of Surjections.** Prove the following statements about surjective functions. (a) If  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$  are surjective, then  $g \circ f$  is surjective. (b) If  $g \circ f$  is surjective, then  $g$  is surjective but  $f$  may not be. (c)  $f : X \rightarrow Y$  is surjective iff for all  $Z$  and all  $g, h : Y \rightarrow Z$ ,  $g \circ f = h \circ f$  implies  $g = h$ .

**1.146. Sorting by Comparisons.** Consider a game in which player 1 picks a permutation  $w$  of  $n$  letters, and player 2 must determine  $w$  by asking player 1 a sequence of yes/no questions. (Player 2 can choose later questions in the sequence based on the answers to earlier questions.) Let  $K(n)$  be the minimum number such that, no matter what  $w$  player 1 chooses, player 2 can correctly identify  $w$  after at most  $K(n)$  questions. (a) Prove that  $(n/2) \log_2(n/2) \leq \lceil \log_2(n!) \rceil \leq K(n)$ . (b) Prove that  $K(n) = \lceil \log_2(n!) \rceil$  for  $n \leq 5$ . (c) Prove that (b) still holds if we restrict player 2 to ask only questions of the form “is  $w_i < w_j$ ?” at each stage. (d) What does (a) imply about the length of time needed to sort  $n$  distinct elements using an algorithm that makes decisions by comparing two data elements at a time?

**1.147.** (a) You are given twelve seemingly identical coins and a balance scale. One coin is counterfeit and is either lighter or heavier than the others. Describe a strategy that can be used to identify which coin is fake in only three weighings. (b) If there are thirteen coins, can the fake coin always be found in three weighings? Justify your answer. (c) If there are  $N$  coins (one of which is fake), derive a lower bound for the number of weighings required to find the fake coin.

**1.148.** Define  $f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}^+$  by  $f(a, b) = 2^a(2b + 1)$ . Prove that  $f$  is a bijection.

**1.149.** Define  $f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$  by  $f(a, b) = ((a + b)^2 + 3a + b)/2$ . Prove that  $f$  is a bijection.

**1.150. Chinese Remainder Theorem.** In this exercise, we write “ $a \bmod k$ ” to denote the unique integer  $b$  in the range  $\{1, 2, \dots, k\}$  such that  $k$  divides  $(a - b)$ . Suppose  $m$  and  $n$  are fixed positive integers. Define a map

$$f : \{1, 2, \dots, mn\} \rightarrow \{1, 2, \dots, m\} \times \{1, 2, \dots, n\} \text{ by setting } f(z) = (z \bmod m, z \bmod n).$$

(a) Show that  $f(z) = f(w)$  iff  $\text{lcm}(m, n)$  divides  $z - w$ . (b) Show that  $f$  is injective iff  $\text{gcd}(m, n) = 1$ . (c) Deduce that  $f$  is a bijection iff  $\text{gcd}(m, n) = 1$ . (d) Prove that for  $\text{gcd}(m, n) = 1$ ,  $f$  maps  $\Phi(mn)$  bijectively onto  $\Phi(m) \times \Phi(n)$ , and hence  $\phi(mn) = \phi(m)\phi(n)$ . (See 1.112 for the definition of  $\Phi$  and  $\phi$ .) (e) Suppose  $n$  has prime factorization  $p_1^{e_1} \cdots p_k^{e_k}$ . Prove that  $\phi(n) = n \prod_{i=1}^k (1 - 1/p_i)$ .

**1.151. Bijective Product Rule.** For any positive integers  $m, n$ , define

$$g : \{0, 1, \dots, m-1\} \times \{0, 1, \dots, n-1\} \rightarrow \{0, 1, \dots, mn-1\}$$

by setting  $g(i, j) = ni + j$ . Carefully prove that  $g$  is a bijection.

**1.152. Bijective Laws of Algebra.** (a) For all sets  $X, Y, Z$ , prove that  $X \cup Y = Y \cup X$ ,  $(X \cup Y) \cup Z = X \cup (Y \cup Z)$ , and  $X \cup \emptyset = X = \emptyset \cup X$ . (b) For all sets  $X, Y, Z$ , define bijections  $f : X \times Y \rightarrow Y \times X$ ,  $g : (X \times Y) \times Z \rightarrow X \times (Y \times Z)$ , and (for  $Y, Z$  disjoint)  $h : X \times (Y \cup Z) \rightarrow (X \times Y) \cup (X \times Z)$ . (c) Use (a), (b), and counting rules to deduce the algebraic laws  $x + y = y + x$ ,  $(x + y) + z = x + (y + z)$ ,  $x + 0 = x = 0 + x$ ,  $xy = yx$ ,  $(xy)z = x(yz)$ , and  $x(y + z) = xy + xz$ , valid for all integers  $x, y, z \geq 0$ .

**1.153. Bijective Laws of Exponents.** (a) If  $X, Y, Z$  are sets with  $Y \cap Z = \emptyset$ , define a bijection from  ${}^{Y \cup Z}X$  to  ${}^Y X \times {}^Z X$ . (b) If  $X, Y, Z$  are any sets, define a bijection from  ${}^Z({}^Y X)$  to  ${}^{Y \times Z} X$ . (c) By specializing to finite sets, deduce the laws of exponents  $x^{y+z} = x^y x^z$  and  $(x^y)^z = x^{yz}$  for all integers  $x, y, z \geq 0$ .

**1.154.** Let  $X$  be any set (possibly infinite). Prove that there exists an injection  $g : X \rightarrow \mathcal{P}(X)$ , but there exists no surjection  $f : X \rightarrow \mathcal{P}(X)$ . Conclude that  $|X| < |\mathcal{P}(X)|$ , and in particular  $n < 2^n$  for all  $n \geq 0$ .

**1.155.** Show that the set of functions  $\mathbb{N}\{0,1\}$  (which can be viewed as the set of infinite sequences of zeroes and ones) is uncountably infinite.

**1.156.** Suppose  $X$  and  $Y$  are sets (possibly infinite),  $f : X \rightarrow Y$  is any function, and  $g : Y \rightarrow X$  is an injective function. (a) Show that there exist sets  $A, B, C, D$  such that  $X$  is the disjoint union of  $A$  and  $B$ ,  $Y$  is the disjoint union of  $C$  and  $D$ ,  $C = f[A] = \{f(x) : x \in A\}$ , and  $B = g[D] = \{g(y) : y \in D\}$ . (Let  $Z = X \sim g[Y]$  and  $h = g \circ f$ ; then let  $A$  be the intersection of all subsets  $U$  of  $X$  such that  $Z \cup h[U] \subseteq U$ .) (b) Deduce the Schröder-Bernstein Theorem from (a).

**1.157.** A sample space  $S$  consists of 25 equally likely outcomes. Suppose we randomly choose an ordered pair  $(A, B)$  of events in  $S$ . (a) Find the probability that  $A$  and  $B$  are disjoint. (b) Find the probability that  $A$  and  $B$  are independent events.

---

## Notes

General treatments of combinatorics may be found in the textbooks [1, 10, 13, 16, 21, 23, 26, 60, 113, 115, 127, 131, 134]. For elementary accounts of probability theory, see [68, 93]. Two advanced probability texts that include measure theory are [11, 30]. More information on the theory of cardinality for infinite sets may be found in [66, 95].

This page intentionally left blank

---

## Combinatorial Identities and Recursions

---

This chapter begins with a discussion of the generalized distributive law and its consequences, which include the multinomial and binomial theorems. We then study algebraic and combinatorial proofs of identities involving binomial coefficients, factorials, summations, etc. We also introduce *recursions*, which provide ways to enumerate classes of combinatorial objects whose cardinalities are not given by closed formulas. We use recursions to obtain information about more intricate combinatorial objects including set partitions, integer partitions, equivalence relations, surjections, and lattice paths.

---

### 2.1 Generalized Distributive Law

Suppose we have a product of several factors, where each factor consists of a sum of some terms. How can we simplify such a product of sums? The following example suggests the answer.

**2.1. Example.** Suppose  $A, B, C, T, U, V$  are  $n \times n$  matrices. Let us simplify the matrix product

$$(C + V)(A + U)(B + C + T).$$

Using the distributive laws for matrices several times, we first compute

$$(C + V)(A + U) = C(A + U) + V(A + U) = CA + CU + VA + VU.$$

Now we multiply this on the right by the matrix  $B + C + T$ . Using the distributive laws again, we obtain

$$\begin{aligned} & CA(B + C + T) + CU(B + C + T) + VA(B + C + T) + VU(B + C + T) \\ &= CAB + CAC + CAT + CUB + CUC + CUT + VAB + VAC + VAT + VUB + VUC + VUT. \end{aligned}$$

Observe that the final answer is a sum of many terms, where each term can be viewed as a *word* drawn from the set of words  $\{C, V\} \times \{A, U\} \times \{B, C, T\}$ . We obtain such a word by choosing a first matrix from the first factor  $C + V$ , then choosing a second matrix from the second factor  $A + U$ , and then choosing a third matrix from the third factor  $B + C + T$ . This sequence of choices can be done in 12 ways, and accordingly there are 12 terms in the final sum.

The pattern in the previous example holds in general. Intuitively, to multiply together some factors, each of which is a sum of some terms, we choose one term from each factor and multiply these terms together. Then we add together all possible products obtained in this way. We will now give a rigorous proof of this result, which ultimately follows from the distributive laws for a (possibly non-commutative) ring. For convenience, we now state the relevant definitions from abstract algebra. Readers unfamiliar with abstract algebra may



replace the abstract rings used below by the set of  $n \times n$  matrices with real entries (which is a particular example of a ring).

**2.2. Definition: Rings.** A *ring* consists of a set  $R$  and two binary operations  $+$  (addition) and  $\cdot$  (multiplication) with domain  $R \times R$ , subject to the following axioms.

$\forall x, y \in R, x + y \in R$	(closure under addition)
$\forall x, y, z \in R, x + (y + z) = (x + y) + z$	(associativity of addition)
$\forall x, y \in R, x + y = y + x$	(commutativity of addition)
$\exists 0_R \in R, \forall x \in R, x + 0_R = x = 0_R + x$	(existence of additive identity)
$\forall x \in R, \exists -x \in R, x + (-x) = 0_R = (-x) + x$	(existence of additive inverses)
$\forall x, y \in R, x \cdot y \in R$	(closure under multiplication)
$\forall x, y, z \in R, x \cdot (y \cdot z) = (x \cdot y) \cdot z$	(associativity of multiplication)
$\exists 1_R \in R, \forall x \in R, x \cdot 1_R = x = 1_R \cdot x$	(existence of multiplicative identity)
$\forall x, y, z \in R, x \cdot (y + z) = x \cdot y + x \cdot z$	(left distributive law)
$\forall x, y, z \in R, (x + y) \cdot z = x \cdot z + y \cdot z$	(right distributive law)

We often write  $xy$  instead of  $x \cdot y$ .  $R$  is a *commutative ring* iff  $R$  satisfies the additional axiom

$$\forall x, y \in R, xy = yx \quad (\text{commutativity of multiplication}).$$

**2.3. Definition: Fields.** A *field* is a commutative ring  $F$  with  $1_F \neq 0_F$  such that every nonzero element of  $F$  has a multiplicative inverse:

$$\forall x \in F, x \neq 0_F \Rightarrow \exists y \in F, xy = 1_F = yx.$$

Let  $R$  be a ring, and suppose  $x_1, x_2, \dots, x_n \in R$ . Because addition is associative, we can unambiguously write a sum like  $x_1 + x_2 + x_3 + \dots + x_n$  without parentheses (see 2.148). Similarly, associativity of multiplication implies that we can write the product  $x_1 x_2 \dots x_n$  without parentheses. Because addition in the ring is commutative, we can permute the summands in a sum like  $x_1 + x_2 + \dots + x_n$  without changing the answer. More formally, for any bijection  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ , we have

$$x_{f(1)} + x_{f(2)} + \dots + x_{f(n)} = x_1 + x_2 + \dots + x_n \in R$$

(see 2.149). It follows that if  $\{x_i : i \in I\}$  is a finite indexed family of ring elements, then the sum of all these elements (denoted  $\sum_{i \in I} x_i$ ) is well defined. Similarly, if  $A$  is a finite subset of  $R$ , then  $\sum_{x \in A} x$  is well defined. On the other hand, the products  $\prod_{i \in I} x_i$  and  $\prod_{x \in A} x$  are *not* well defined (for  $R$  non-commutative) unless we specify in advance a total ordering on  $I$  and  $A$ .

Now we are ready to derive the general distributive law for non-commutative rings. The idea is to keep iterating the left and right distributive laws to obtain successively more general formulas. We divide the proof into a sequence of lemmas.

**2.4. Lemma.** Let  $R$  be a ring,  $n$  a positive integer, and  $x, y_1, y_2, \dots, y_n \in R$ . Then

$$x(y_1 + y_2 + \dots + y_n) = xy_1 + xy_2 + \dots + xy_n; \quad (y_1 + y_2 + \dots + y_n)x = y_1x + y_2x + \dots + y_nx.$$

*Proof.* We prove the first equation by induction on  $n$ . The case  $n = 1$  is immediate, while the case  $n = 2$  is the left distributive law for  $R$ . Now assume that  $x(y_1 + y_2 + \dots + y_n) = xy_1 + xy_2 + \dots + xy_n$  is known for some  $n \geq 2$ ; let us prove the corresponding formula for  $n + 1$ . In the left distributive law for  $R$ , let  $y = y_1 + \dots + y_n$  and  $z = y_{n+1}$ . Using this and the induction hypothesis, we calculate

$$x(y_1 + \dots + y_n + y_{n+1}) = x(y + z) = xy + xz = x(y_1 + \dots + y_n) + xy_{n+1} = xy_1 + \dots + xy_n + xy_{n+1}.$$

This proves the first equation. The second is proved similarly from the right distributive law.  $\square$

Now suppose  $x \in R$  and  $\{y_i : i \in I\}$  is a finite indexed family of elements of  $R$ . The results in the previous lemma can be written as follows:

$$x \cdot \left( \sum_{i \in I} y_i \right) = \sum_{i \in I} (x \cdot y_i); \quad (2.1)$$

$$\left( \sum_{i \in I} y_i \right) \cdot x = \sum_{i \in I} (y_i \cdot x). \quad (2.2)$$

**2.5. Lemma.** Let  $R$  be a ring, and suppose  $\{u_i : i \in I\}$  and  $\{v_j : j \in J\}$  are two finite indexed families of ring elements. Then

$$\left( \sum_{i \in I} u_i \right) \cdot \left( \sum_{j \in J} v_j \right) = \sum_{(i,j) \in I \times J} (u_i \cdot v_j). \quad (2.3)$$

*Proof.* Applying (2.2) with  $x = \sum_{j \in J} v_j \in R$  and  $y_i = u_i$  for all  $i$ , we obtain first

$$\left( \sum_{i \in I} u_i \right) \cdot \left( \sum_{j \in J} v_j \right) = \sum_{i \in I} \left( u_i \cdot \sum_{j \in J} v_j \right).$$

Now, for each  $i \in I$ , apply (2.1) with  $x = u_i$  and  $\sum_{i \in I} y_i$  replaced by  $\sum_{j \in J} v_j$  to obtain

$$\sum_{i \in I} \left( u_i \cdot \sum_{j \in J} v_j \right) = \sum_{i \in I} \left( \sum_{j \in J} (u_i \cdot v_j) \right).$$

Finally, since addition in  $R$  is commutative, the iterated sum in the last formula is equal to the single sum

$$\sum_{(i,j) \in I \times J} (u_i \cdot v_j)$$

over the new index set  $I \times J$ . The lemma follows.  $\square$

**2.6. Theorem (Generalized Distributive Law).** Suppose  $R$  is a ring,  $I_1, \dots, I_n$  are finite index sets, and  $\{x_{k,i_k} : i_k \in I_k\}$  are indexed families of ring elements for  $1 \leq k \leq n$ . Then

$$\left( \sum_{i_1 \in I_1} x_{1,i_1} \right) \cdot \left( \sum_{i_2 \in I_2} x_{2,i_2} \right) \cdot \dots \cdot \left( \sum_{i_n \in I_n} x_{n,i_n} \right) = \sum_{(i_1, \dots, i_n) \in I_1 \times \dots \times I_n} (x_{1,i_1} \cdot x_{2,i_2} \cdot \dots \cdot x_{n,i_n}). \quad (2.4)$$

We can also write this as

$$\prod_{k=1}^n \left( \sum_{i_k \in I_k} x_{k,i_k} \right) = \sum_{(i_1, \dots, i_n) \in I_1 \times \dots \times I_n} \left( \prod_{k=1}^n x_{k,i_k} \right), \quad (2.5)$$

provided we remember that the *order* of the factors in  $\prod_{k=1}^n u_k = u_1 u_2 \dots u_n$  is crucial.

*Proof.* We use induction on  $n$ . There is nothing to prove if  $n = 1$ , and the case  $n = 2$  was proved in the previous lemma. Now assume the result holds for some  $n \geq 2$ ; we prove the

result for  $n + 1$  factors. Using  $\prod_{k=1}^{n+1} v_k = (\prod_{k=1}^n v_k) \cdot v_{n+1}$ , then the induction hypothesis for  $n$  factors, then the result for 2 factors, we compute

$$\begin{aligned} \prod_{k=1}^{n+1} \sum_{i_k \in I_k} x_{k,i_k} &= \left( \prod_{k=1}^n \sum_{i_k \in I_k} x_{k,i_k} \right) \cdot \left( \sum_{i_{n+1} \in I_{n+1}} x_{n+1,i_{n+1}} \right) \\ &= \left( \sum_{(i_1, \dots, i_n) \in I_1 \times \dots \times I_n} \prod_{k=1}^n x_{k,i_k} \right) \cdot \left( \sum_{i_{n+1} \in I_{n+1}} x_{n+1,i_{n+1}} \right) \\ &= \sum_{((i_1, \dots, i_n), i_{n+1}) \in (I_1 \times \dots \times I_n) \times I_{n+1}} \left( \left( \prod_{k=1}^n x_{k,i_k} \right) \cdot x_{n+1,i_{n+1}} \right). \end{aligned}$$

By commutativity of addition, the final expression is equal to

$$\sum_{(i_1, \dots, i_{n+1}) \in I_1 \times \dots \times I_{n+1}} \prod_{k=1}^{n+1} x_{k,i_k}.$$

This completes the induction.  $\square$

Here is a formula that follows from the generalized distributive law.

**2.7. Theorem.** If  $y_1, \dots, y_n, z_1, \dots, z_n$  are elements of a *commutative* ring  $R$ , then

$$\prod_{k=1}^n (y_k + z_k) = \sum_{S \subseteq \{1, 2, \dots, n\}} \prod_{k \in S} z_k \prod_{k \notin S} y_k.$$

*Proof.* Write  $x_{k,0} = y_k$  and  $x_{k,1} = z_k$  for  $1 \leq k \leq n$ , and let  $I_1 = I_2 = \dots = I_n = \{0, 1\}$ . Using 2.6 gives

$$\prod_{k=1}^n (y_k + z_k) = \prod_{k=1}^n (x_{k,0} + x_{k,1}) = \sum_{(i_1, \dots, i_n) \in \{0, 1\}^n} \prod_{k=1}^n x_{k,i_k}.$$

Now, we can use the bijection in 1.38 to convert the sum over binary words in  $\{0, 1\}^n$  to a sum over subsets  $S \subseteq \{1, 2, \dots, n\}$ . Suppose  $(i_1, \dots, i_n)$  corresponds to  $S$  under the bijection. Then  $k \in S$  iff  $i_k = 1$  iff  $x_{k,i_k} = z_k$ , while  $k \notin S$  iff  $i_k = 0$  iff  $x_{k,i_k} = y_k$ . It follows that the summand indexed by  $S$  is

$$\prod_{k=1}^n x_{k,i_k} = \prod_{k:i_k=1} x_{k,i_k} \prod_{k:i_k=0} x_{k,i_k} = \prod_{k \in S} z_k \prod_{k \notin S} y_k.$$

Note that the first equality here used the commutativity of  $R$ .  $\square$

## 2.2 Multinomial and Binomial Theorems

We now deduce some consequences of the generalized distributive law. In particular, we derive the non-commutative and commutative versions of the multinomial theorem and the binomial theorem.

**2.8. Theorem.** Suppose  $R$  is a ring and  $A_1, \dots, A_n$  are finite subsets of  $R$ . Then

$$\left( \sum_{w_1 \in A_1} w_1 \right) \cdot \left( \sum_{w_2 \in A_2} w_2 \right) \cdot \dots \cdot \left( \sum_{w_n \in A_n} w_n \right) = \sum_{(w_1, w_2, \dots, w_n) \in A_1 \times A_2 \times \dots \times A_n} w_1 w_2 \cdots w_n.$$

*Proof.* In 2.6, choose the index sets  $I_k = A_k$ , and define  $x_{k, i_k} = i_k \in A_k \subseteq R$  for each  $k \leq n$  and each  $i_k \in I_k$ . Then  $\sum_{w_k \in I_k} x_{k, w_k} = \sum_{w_k \in A_k} w_k$  for each  $k$ , and  $\prod_{k=1}^n x_{k, w_k} = w_1 w_2 \cdots w_n$ . Thus the formula in the theorem is a special case of (2.4).  $\square$

To emphasize the combinatorial nature of the previous result, we can write it as follows:

$$\left( \sum_{w_1 \in A_1} w_1 \right) \cdot \left( \sum_{w_2 \in A_2} w_2 \right) \cdot \dots \cdot \left( \sum_{w_n \in A_n} w_n \right) = \sum_{\text{words } w \in A_1 \times A_2 \times \dots \times A_n} w.$$

Intuitively, we simplify a given product of sums by choosing one letter  $w_i$  from each factor, concatenating (multiplying) these letters to get a word, and adding all the words obtainable in this way.

**2.9. Non-Commutative Multinomial Theorem.** Suppose  $R$  is a ring,  $n \in \mathbb{N}$ , and  $A = \{z_1, \dots, z_s\} \subseteq R$ . Then

$$(z_1 + z_2 + \dots + z_s)^n = \sum_{w \in A^n} w_1 w_2 \cdots w_n.$$

*Proof.* Take  $A_1 = A_2 = \dots = A_n = A$  in 2.8.  $\square$

**2.10. Example.** If  $A$  and  $B$  are two  $n \times n$  matrices, then

$$(A + B)^3 = AAA + AAB + ABA + ABB + BAA + BAB + BBA + BBB.$$

If  $A, B, \dots, Z$  are 26 matrices, then

$$\begin{aligned} (A + B + \dots + Z)^4 &= AAAA + AAAB + AABA + \dots + ZZZY + ZZZZ \\ &= \text{the sum of all 4-letter words.} \end{aligned}$$

**2.11. Remark.** Our statement of the non-commutative multinomial theorem tacitly assumed that  $z_1, \dots, z_s$  were *distinct* ring elements. This assumption can be dropped at the expense of a slight notation change. More precisely, if  $\{z_i : i \in I\}$  is a finite indexed family of ring elements, then it follows from 2.6 that

$$\left( \sum_{i \in I} z_i \right)^n = \sum_{w \in I^n} z_{w_1} z_{w_2} \cdots z_{w_n}.$$

Similar comments apply to the theorems below.

**2.12. Commutative Multinomial Theorem.** Suppose  $R$  is a ring,  $n \in \mathbb{N}$ , and  $z_1, \dots, z_s \in R$  are elements of  $R$  that *commute* (meaning  $z_i z_j = z_j z_i$  for all  $i, j$ ). Then

$$(z_1 + z_2 + \dots + z_s)^n = \sum_{n_1 + n_2 + \dots + n_s = n} \binom{n}{n_1, n_2, \dots, n_s} z_1^{n_1} z_2^{n_2} \cdots z_s^{n_s}.$$

The summation here extends over all ordered sequences  $(n_1, n_2, \dots, n_s)$  of nonnegative integers that sum to  $n$ .

*Proof.* Let  $A = \{z_1, z_2, \dots, z_s\} \subseteq R$ . Let  $X$  be the set of all ordered sequences  $(n_1, n_2, \dots, n_s)$  of nonnegative integers that sum to  $n$ . The non-commutative multinomial theorem gives

$$(z_1 + z_2 + \dots + z_s)^n = \sum_{w \in A^n} w_1 w_2 \dots w_n.$$

The set  $A^n$  of  $n$ -letter words over the alphabet  $A$  is the disjoint union of the sets  $\mathcal{R}(z_1^{n_1} z_2^{n_2} \dots z_s^{n_s})$ , as  $(n_1, \dots, n_s)$  ranges over  $X$ . By commutativity of addition, we therefore have

$$(z_1 + z_2 + \dots + z_s)^n = \sum_{(n_1, \dots, n_s) \in X} \sum_{w \in \mathcal{R}(z_1^{n_1} \dots z_s^{n_s})} w_1 w_2 \dots w_n.$$

Now we use the assumption that all the  $z_i$ 's commute with one another. This assumption allows us to reorder any product of  $z_i$ 's so that all  $z_1$ 's come first, followed by the  $z_2$ 's, etc. Given  $w \in \mathcal{R}(z_1^{n_1} \dots z_s^{n_s})$ , reordering the letters of  $w$  gives

$$w_1 w_2 \dots w_n = z_1^{n_1} z_2^{n_2} \dots z_s^{n_s}.$$

Thus,

$$\begin{aligned} (z_1 + z_2 + \dots + z_s)^n &= \sum_{(n_1, \dots, n_s) \in X} \sum_{w \in \mathcal{R}(z_1^{n_1} \dots z_s^{n_s})} z_1^{n_1} \dots z_s^{n_s} \\ &= \sum_{(n_1, \dots, n_s) \in X} z_1^{n_1} \dots z_s^{n_s} \cdot \left( \sum_{w \in \mathcal{R}(z_1^{n_1} \dots z_s^{n_s})} 1_R \right). \end{aligned}$$

The inner sum is  $|\mathcal{R}(z_1^{n_1} \dots z_s^{n_s})| = \binom{n}{n_1, n_2, \dots, n_s}$  (or more precisely, the sum of this many copies of  $1_R$ ). Thus, we obtain the formula in the statement of the theorem.  $\square$

**2.13. Example.** If  $xy = yx$  and  $xz = zx$  and  $yz = zy$ , then

$$(x + y + z)^3 = x^3 + y^3 + z^3 + 3x^2y + 3x^2z + 3y^2z + 3xy^2 + 3xz^2 + 3yz^2 + 6xyz.$$

**2.14. Commutative Binomial Theorem.** Suppose  $R$  is a ring,  $n \in \mathbb{N}$ , and  $x, y \in R$  are ring elements such that  $xy = yx$ . Then

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$

*Proof.* Apply the commutative multinomial theorem with  $s = 2$ ,  $z_1 = x$ , and  $z_2 = y$  to get

$$(x + y)^n = \sum_{n_1 + n_2 = n} \binom{n}{n_1, n_2} x^{n_1} y^{n_2}.$$

Let  $n_1 = k$ ; note that the possible values of  $k$  are  $0, 1, \dots, n$ . Once  $n_1$  has been chosen,  $n_2$  is uniquely determined as  $n_2 = n - n_1 = n - k$ . Also,  $\binom{n}{n_1, n_2} = \binom{n}{k}$ , so the formula becomes

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}. \quad \square$$

**2.15. Example.** What is the coefficient of  $z^7$  in  $(2z - 5)^9$ ? We apply the binomial theorem taking  $x = 2z$  and  $y = -5$  and  $n = 9$ . We have

$$(2z - 5)^9 = \sum_{k=0}^9 \binom{9}{k} (2z)^k (-5)^{9-k}.$$

The only summand involving  $z^7$  is the  $k = 7$  summand. The corresponding coefficient is

$$\binom{9}{7} 2^7 (-5)^2 = 115,200.$$

**2.16. Remark.** If  $r$  is any real number and  $x$  is a real number such that  $|x| < 1$ , there exists a power series expansion for  $(1 + x)^r$  that is often called the *generalized binomial formula*. This power series is discussed in 7.68.

## 2.3 Combinatorial Proofs

Consider the problem of proving an identity of the form  $A = B$ , where  $A$  and  $B$  are formulas that may involve factorials, binomial coefficients, powers, etc. One way to prove such an identity is to give an *algebraic proof* using tools like the binomial theorem or other algebraic techniques. Another way to prove such an identity is to find a *combinatorial proof*. A combinatorial proof establishes the equality of two formulas by exhibiting a set of objects whose cardinality is given by both formulas. Thus, the main steps in a combinatorial proof of  $A = B$  are as follows.

- Define a set  $S$  of objects.
- Give a counting argument (using the sum rule, product rule, bijections, etc.) to prove that  $|S| = A$ .
- Give a different counting argument to prove that  $|S| = B$ .
- Conclude that  $A = B$ .

We now give some examples illustrating this technique and its variations.

**2.17. Theorem.** For all  $n \in \mathbb{N}$ ,

$$2^n = \sum_{k=0}^n \binom{n}{k}.$$

*Proof.* We give an algebraic proof and a combinatorial proof.

*Algebraic Proof.* By the binomial theorem, we know that

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \quad (x, y \in \mathbb{R}).$$

Setting  $x = y = 1$  yields the desired formula.

*Combinatorial Proof.* Fix  $n \in \mathbb{N}$ . Let  $S$  be the set of all subsets of  $\{1, 2, \dots, n\}$ . As shown

earlier,  $|S| = 2^n$  since we can build a typical subset by either including or excluding each of the  $n$  available elements. On the other hand, note that  $S$  is the disjoint union

$$S = S_0 \cup S_1 \cup \cdots \cup S_n,$$

where  $S_k$  consists of all  $k$ -element subsets of  $\{1, 2, \dots, n\}$ . As shown earlier,  $|S_k| = \binom{n}{k}$ . By the sum rule, we therefore have

$$|S| = \sum_{k=0}^n |S_k| = \sum_{k=0}^n \binom{n}{k}.$$

Thus,

$$2^n = |S| = \sum_{k=0}^n \binom{n}{k}.$$

□

**2.18. Theorem.** For all integers  $n, k$  with  $0 \leq k \leq n$ , we have

$$\binom{n}{k} = \binom{n}{n-k}.$$

*Proof.* Again we give both an algebraic proof and a combinatorial proof.

*Algebraic Proof.* Using the explicit formula for binomial coefficients involving factorials, we calculate

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!(n-(n-k))!} = \binom{n}{n-k}.$$

*Combinatorial Proof.* For this proof, we will define two different sets of objects. Fix  $n$  and  $k$ . Let  $S$  be the set of all  $k$ -element subsets of  $\{1, 2, \dots, n\}$ , and let  $T$  be the set of all  $(n-k)$ -element subsets of  $\{1, 2, \dots, n\}$ . We have already shown that  $|S| = \binom{n}{k}$  and  $|T| = \binom{n}{n-k}$ . We complete the proof by exhibiting a bijection  $\phi : S \rightarrow T$ , which shows that  $|S| = |T|$ . Given  $A \in S$ , define  $\phi(A) = \{1, 2, \dots, n\} \setminus A$ . Since  $A$  has  $k$  elements,  $\phi(A)$  has  $n-k$  elements and is thus an element of  $T$ . The inverse of this map is the map  $\phi' : T \rightarrow S$  given by  $\phi'(B) = \{1, 2, \dots, n\} \setminus B$  for  $B \in T$ . Note that  $\phi$  and  $\phi'$  are both restrictions of the “set complement” map  $I : \mathcal{P}(\{1, 2, \dots, n\}) \rightarrow \mathcal{P}(\{1, 2, \dots, n\})$  given by  $I(A) = \{1, 2, \dots, n\} \setminus A$ . Since  $I \circ I$  is the identity map on  $\mathcal{P}(\{1, 2, \dots, n\})$ , it follows that  $\phi'$  is the two-sided inverse of  $\phi$ . □

**2.19. Theorem.** For  $0 \leq k \leq n$ ,

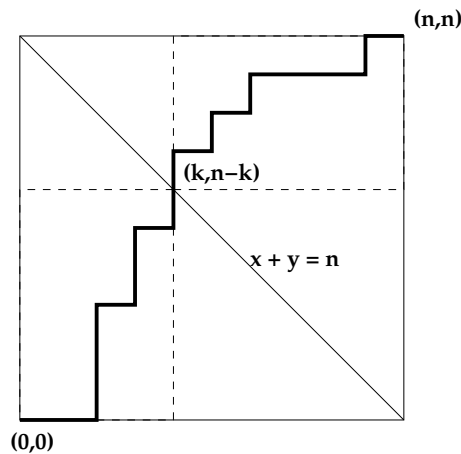
$$\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}.$$

*Proof.* In terms of factorials, we are trying to prove that

$$\sum_{k=0}^n \frac{(n!)^2}{(k!)^2((n-k)!)^2} = \frac{(2n)!}{n!n!}.$$

An algebraic proof of this formula is not evident. So we proceed to look for a combinatorial proof.

Define  $S$  to be the set of all  $n$ -element subsets of  $X = \{1, 2, \dots, 2n\}$ . This choice of  $S$  was motivated by our knowledge that  $|S| = \binom{2n}{n}$ , which is the right side of the desired


**FIGURE 2.1**

A combinatorial proof using lattice paths.

identity. To complete the proof, we count  $S$  in a new way. Let  $X_1 = \{1, 2, \dots, n\}$  and  $X_2 = \{n+1, \dots, 2n\}$ . For  $0 \leq k \leq n$ , define

$$S_k = \{A \in S : |A \cap X_1| = k \text{ and } |A \cap X_2| = n - k\}.$$

Evidently,  $S$  is the disjoint union of the  $S_k$ 's, so that  $|S| = \sum_{k=0}^n |S_k|$  by the sum rule. To compute  $|S_k|$ , we build a typical object  $A \in S_k$  by making two choices. First, choose the  $k$ -element subset  $A \cap X_1$  in any of  $\binom{n}{k}$  ways. Second, choose the  $(n-k)$ -element subset  $A \cap X_2$  in any of  $\binom{n}{n-k} = \binom{n}{k}$  ways. We see that  $|S_k| = \binom{n}{k}^2$  by the product rule. Thus,  $|S| = \sum_{k=0}^n \binom{n}{k}^2$ , completing the proof.  $\square$

One can often find different combinatorial proofs of a given identity. For example, here is an alternate proof of the previous identity using lattice paths. Let  $S$  be the set of all lattice paths from the origin to  $(n, n)$ ; we know that  $|S| = \binom{2n}{n, n} = \binom{2n}{n}$ . For  $0 \leq k \leq n$ , let  $S_k$  be the set of all paths in  $S$  passing through the point  $(k, n-k)$  on the line  $x + y = n$ . Every path in  $S$  must go through exactly one such point for some  $k$  between 0 and  $n$ , so  $S$  is the disjoint union of  $S_0, S_1, \dots, S_n$ . See Figure 2.1. To build a path in  $S_k$ , first choose a path from  $(0, 0)$  to  $(k, n-k)$  in any of  $\binom{n}{k, n-k} = \binom{n}{k}$  ways. Second, choose a path from  $(k, n-k)$  to  $(n, n)$ . This is a path in a rectangle of width  $n-k$  and height  $n - (n-k) = k$ , so there are  $\binom{n}{n-k, k} = \binom{n}{k}$  ways to make this second choice. By the sum and product rules, we conclude that

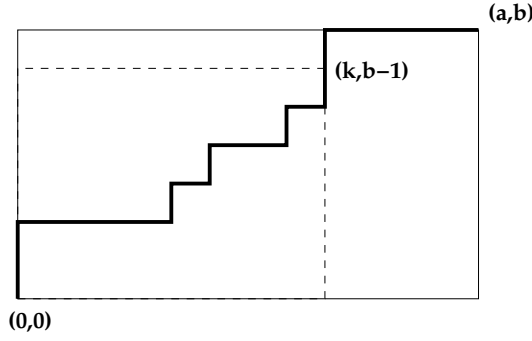
$$|S| = \sum_{k=0}^n |S_k| = \sum_{k=0}^n \binom{n}{k}^2.$$

Lattice paths can often be used to give elegant, visually appealing combinatorial proofs of identities involving binomial coefficients. We conclude this section with two more examples of this kind.

**2.20. Theorem.** For all integers  $a \geq 0$  and  $b \geq 1$ ,

$$\binom{a+b}{a, b} = \sum_{k=0}^a \binom{k+b-1}{k, b-1}.$$



**FIGURE 2.2**

Another combinatorial proof using lattice paths.

*Proof.* Let  $S$  be the set of all lattice paths from the origin to  $(a, b)$ . We already know that  $|S| = \binom{a+b}{a, b}$ . For  $0 \leq k \leq a$ , let  $S_k$  be the set of paths  $\pi \in S$  such that the last north step of  $\pi$  lies on the line  $x = k$ . See Figure 2.2. We can build a path in  $S_k$  by choosing any lattice path from the origin to  $(k, b-1)$  in  $\binom{k+b-1}{k, b-1}$  ways, and then appending one north step and  $a-k$  east steps. Thus, the required identity follows from the sum rule. If we classify the paths by the final east step instead, we obtain the dual identity (for  $a \geq 1, b \geq 0$ )

$$\binom{a+b}{a, b} = \sum_{j=0}^b \binom{a-1+j}{a-1, j}.$$

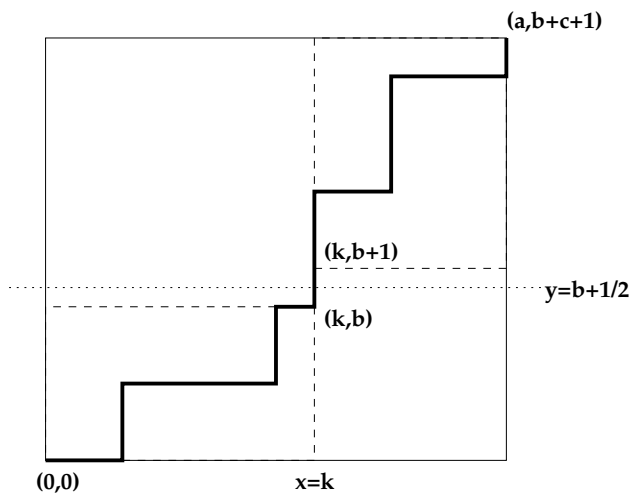
This identity also follows from the previous one by the known symmetry  $\binom{a+b}{a, b} = \binom{b+a}{b, a}$ .  $\square$

**2.21. Theorem (Chu-Vandermonde Identity).** For all integers  $a, b, c \geq 0$ ,

$$\binom{a+b+c+1}{a, b+c+1} = \sum_{k=0}^a \binom{k+b}{k, b} \binom{a-k+c}{a-k, c}.$$

*Proof.* Let  $S$  be the set of all lattice paths from the origin to  $(a, b+c+1)$ . We know that  $|S| = \binom{a+b+c+1}{a, b+c+1}$ . For  $0 \leq k \leq a$ , let  $S_k$  be the set of paths  $\pi \in S$  that contain the north step from  $(k, b)$  to  $(k, b+1)$ . Since every path in  $\pi$  must cross the line  $y = b+1/2$  by taking a north step between the lines  $x = 0$  and  $x = a$ , we see that  $S$  is the disjoint union of  $S_0, S_1, \dots, S_a$ . See Figure 2.3. Now, we can build a path in  $S_k$  as follows. First, choose a lattice path from the origin to  $(k, b)$  in  $\binom{k+b}{k, b}$  ways. Second, append a north step to this path. Third, choose a lattice path from  $(k, b+1)$  to  $(a, b+c+1)$ . This is a path in a rectangle of width  $a-k$  and height  $c$ , so there are  $\binom{a-k+c}{a-k, c}$  ways to make this choice. Thus,  $|S_k| = \binom{k+b}{k, b} \cdot 1 \cdot \binom{a-k+c}{a-k, c}$  by the product rule. The desired identity now follows from the sum rule.  $\square$

We remark that 2.20 is the special case of 2.21 obtained by setting  $c = 0$  and replacing  $b$  by  $b-1$ .

**FIGURE 2.3**

A third combinatorial proof using lattice paths.

## 2.4 Recursions

Suppose we are given some unknown quantities  $a_0, a_1, \dots, a_n, \dots$ . A *closed formula* for these quantities is an expression of the form  $a_n = f(n)$ , where the right side is some explicit formula involving the integer  $n$  but not involving any of the unknown quantities  $a_i$ . In contrast, a *recursive formula* for  $a_n$  is an expression of the form  $a_n = f(n, a_0, a_1, \dots, a_{n-1})$ , where the right side is a formula that does involve one or more of the unknown quantities  $a_i$ . A recursive formula is usually accompanied by one or more *initial conditions*, which are non-recursive expressions for  $a_0$  and possibly other  $a_i$ 's. Similar definitions apply to doubly indexed sequences  $a_{n,k}$ .

Now consider the problem of counting sets of combinatorial objects. Suppose we have several related families of objects, say  $T_0, T_1, \dots, T_n, \dots$ . We think of the index  $n$  as somehow measuring the size of the objects in  $T_n$ . Sometimes we can give an explicit description of the objects in  $T_n$  (using the sum and product rules) leading to a closed formula for  $|T_n|$ . In many cases, however, it is more natural to give a *recursive* description of  $T_n$ , which tells us how to construct a typical object in  $T_n$  by assembling smaller objects of the same kind from the sets  $T_0, \dots, T_{n-1}$ . Such an argument leads to a recursive formula for  $|T_n|$  in terms of one or more of the quantities  $|T_0|, \dots, |T_{n-1}|$ . If we suspect that  $|T_n|$  is also given by a certain closed formula, we can then prove this fact using induction. We use the following example to illustrate these ideas.

**2.22. Theorem: Recursion for Subsets.** For each integer  $n \geq 0$ , let  $T_n$  be the set of all subsets of  $\{1, 2, \dots, n\}$ , and let  $a_n = |T_n|$ . We derive a recursive formula for  $a_n$  as follows. Suppose  $n \geq 1$  and we are trying to build a typical subset  $A \in T_n$ . We can do this recursively by first choosing a subset  $A' \subseteq \{1, 2, \dots, n-1\}$  in any of  $|T_{n-1}| = a_{n-1}$  ways, and then either adding or not adding the element  $n$  to this subset (2 possibilities). By the product rule, we conclude that

$$a_n = a_{n-1} \cdot 2 \quad (n \geq 1).$$

The initial condition is  $a_0 = 1$ , since  $T_0 = \{\emptyset\}$ .

Using the recursion and initial condition, we calculate:

$$(a_0, a_1, a_2, a_3, a_4, a_5, \dots) = (1, 2, 4, 8, 16, 32, \dots).$$

The pattern suggests that  $a_n = 2^n$  for all  $n \geq 0$ . (We have already proved this earlier, but we wish to reprove this fact using our recursion.) We will prove that  $a_n = 2^n$  by induction on  $n$ . In the base case ( $n = 0$ ), we have  $a_0 = 1 = 2^0$  by the initial condition. Assume that  $n > 0$  and that  $a_{n-1} = 2^{n-1}$  (this is the induction hypothesis). Using the recursion and the induction hypothesis, we see that

$$a_n = 2a_{n-1} = 2(2^{n-1}) = 2^n.$$

This completes the proof by induction.

**2.23. Example: Fibonacci Words.** Let  $W_n$  be the set of all words in  $\{0, 1\}^n$  that do not have two consecutive zeroes, and let  $f_n = |W_n|$ . We now derive a recursion and initial condition for the sequence of  $f_n$ 's. First, direct enumeration shows that  $f_0 = 1$  and  $f_1 = 2$ . Suppose  $n \geq 2$ . We use the sum rule to find a formula for  $|W_n| = f_n$ . Given  $w \in W_n$ ,  $w$  starts with either 0 or 1. If  $w_1 = 0$ , then we are forced to have  $w_2 = 1$ , and then  $w' = w_3w_4 \cdots w_n$  can be an arbitrary word in  $W_{n-2}$ . Therefore, there are  $f_{n-2}$  words in  $W_n$  starting with 0. On the other hand, if  $w_1 = 1$ , then  $w' = w_2 \cdots w_n$  can be an arbitrary word in  $W_{n-1}$ . Therefore, there are  $f_{n-1}$  words in  $W_n$  starting with 1. By the sum rule,

$$f_n = f_{n-1} + f_{n-2} \quad (n \geq 2).$$

Using this recursion and the initial conditions, we compute

$$(f_0, f_1, f_2, f_3, f_4, f_5, \dots) = (1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \dots).$$

This sequence is called the *Fibonacci sequence*. We will find an explicit closed formula for  $f_n$  later (see 2.134(a) or §7.14).

Now we consider some examples involving doubly indexed families of combinatorial objects. We begin by revisiting the enumeration of  $k$ -permutations, subsets, multisets, and anagrams. We will reprove some of our earlier counting results by recursive methods.

**2.24. Recursion for  $k$ -Permutations.** For all integers  $n, k \geq 0$ , let  $P(n, k)$  be the number of  $k$ -permutations of an  $n$ -element set. (One can show bijectively that every  $n$ -element set has the same number of  $k$ -permutations, so that  $P(n, k)$  is a well-defined integer. Alternatively, we could define  $P(n, k)$  to be the number of  $k$ -permutations of a particular  $n$ -element set like  $\{1, 2, \dots, n\}$ . However, in the latter case, the argument in the text must be modified accordingly. Similar comments apply to later examples.) Recall that a  $k$ -permutation is an ordered sequence  $x_1x_2 \cdots x_k$  of *distinct* elements from the given  $n$ -element set. Observe that  $P(n, k) = 0$  whenever  $k > n$ . On the other hand,  $P(n, 0) = 1$  for all  $n \geq 0$  since the empty sequence is the unique 0-permutation of any  $n$ -element set. Now assume that  $0 < k \leq n$ . We can build a typical  $k$ -permutation  $x = x_1x_2 \cdots x_k$  of a given  $n$ -element set  $X$  as follows. First, choose  $x_1$  in any of  $n$  ways. For the second choice, note that  $x' = x_2x_3 \cdots x_k$  can be any  $(k-1)$ -permutation of the  $(n-1)$ -element set  $X \sim \{x_1\}$ . There are  $P(n-1, k-1)$  choices for  $x'$ , by definition. The product rule thus gives us the recursion

$$P(n, k) = nP(n-1, k-1) \quad (0 < k \leq n).$$

The initial conditions are  $P(n, 0) = 1$  for all  $n$  and  $P(n, k) = 0$  for all  $k > n$ .

In §1.4, we used the product rule to prove that  $P(n, k) = n(n-1) \cdots (n-k+1) =$

$n!/(n-k)!$  for  $0 \leq k \leq n$ . Let us now reprove this result using our recursion. We proceed by induction on  $n$ . In the base case,  $n = 0$  and hence  $k = 0$ . The initial condition gives

$$P(n, k) = P(0, 0) = 1 = 0!/(0-0)! = n!/(n-k)!.$$

For the induction step, assume that  $n > 0$  and that

$$P(n-1, j) = (n-1)!/(n-1-j)! \quad (0 \leq j \leq n-1).$$

Fix  $k$  with  $0 \leq k \leq n$ . If  $k = 0$ , the initial condition gives

$$P(n, k) = P(n, 0) = 1 = n!/(n-0)! = n!/(n-k)!.$$

If  $k > 0$ , we use the recursion and induction hypothesis (applied to  $j = k-1$ ) to compute

$$P(n, k) = nP(n-1, k-1) = n \frac{(n-1)!}{((n-1)-(k-1))!} = \frac{n!}{(n-k)!}.$$

This completes the proof by induction.

**2.25. Recursion for  $k$ -element Subsets.** For all integers  $n, k \geq 0$ , let  $C(n, k)$  be the number of  $k$ -element subsets of  $\{1, 2, \dots, n\}$ . Observe that  $C(n, k) = 0$  whenever  $k > n$ . On the other hand, the initial condition  $C(n, 0) = 1$  follows since the empty set is the unique zero-element subset of any set. Similarly,  $C(n, n) = 1$  since  $\{1, 2, \dots, n\}$  is the only  $n$ -element subset of itself. Let us now derive a recursion for  $C(n, k)$  assuming that  $0 < k < n$ . A typical  $k$ -element subset  $A$  of  $\{1, 2, \dots, n\}$  either *does* or *does not* contain  $n$  as a member. In the former case, we can construct  $A$  by choosing any  $(k-1)$ -element subset of  $\{1, 2, \dots, n-1\}$  in  $C(n-1, k-1)$  ways, and then appending  $n$  as the final member of  $A$ . In the latter case, we can construct  $A$  by choosing any  $k$ -element subset of  $\{1, 2, \dots, n-1\}$  in  $C(n-1, k)$  ways. By the sum rule, we deduce *Pascal's recursion*

$$C(n, k) = C(n-1, k-1) + C(n-1, k) \quad (0 < k < n).$$

For  $n > 0$ , this recursion even holds for  $k = 0$  and  $k = n$ , provided we use the conventions that  $C(a, b) = 0$  whenever  $b < 0$  or  $b > a$ .

In §1.8, we proved that  $C(n, k) = \frac{n!}{k!(n-k)!} = \binom{n}{k}$ . Let us reprove this result using the recursion and initial conditions. We proceed by induction on  $n$ . The base case  $n = k = 0$  follows since  $C(0, 0) = 1 = \frac{0!}{0!(0-0)!}$ . Assume  $n > 0$  and that

$$C(n-1, j) = \frac{(n-1)!}{j!(n-1-j)!} \quad (0 \leq j \leq n-1).$$

Fix  $k$  with  $0 \leq k \leq n$ . If  $k = 0$ , the initial condition gives

$$C(n, k) = 1 = \frac{n!}{0!(n-0)!}$$

as desired. Similarly, the result holds when  $k = n$ . If  $0 < k < n$ , we use the recursion and induction hypothesis (applied to  $j = k-1$  and to  $j = k$ , which are integers between 0 and



studying combinatorial collections whose cardinalities may not be given by explicit closed formulas. Nevertheless, these cardinalities satisfy recursions that allow them to be computed quickly and efficiently.

## 2.5 Recursions for Multisets and Anagrams

This section continues to give examples of combinatorial recursions for objects we have studied before, namely multisets and anagrams.

**2.26. Recursion for Multisets.** In §1.11, we counted  $k$ -element multisets on an  $n$ -letter alphabet using bijective techniques. Now, we give a recursive analysis to reprove the enumeration results for multisets. For all integers  $n, k \geq 0$ , let  $M(n, k)$  be the number of  $k$ -element multisets using letters from  $\{1, 2, \dots, n\}$ . The initial conditions are  $M(n, 0) = 1$  for all  $n \geq 0$  and  $M(0, k) = 0$  for all  $k > 0$ . We now derive a recursion for  $M(n, k)$  assuming  $n > 0$  and  $k > 0$ . A typical multiset counted by  $M(n, k)$  either does not contain  $n$  at all or contains one or more copies of  $n$ . In the former case, the multiset is a  $k$ -element multiset using letters from  $\{1, 2, \dots, n-1\}$ , and there are  $M(n-1, k)$  such multisets. In the latter case, if we remove one copy of  $n$  from the multiset, we obtain an arbitrary  $(k-1)$ -element multiset using letters from  $\{1, 2, \dots, n\}$ . There are  $M(n, k-1)$  such multisets. By the sum rule, we obtain the recursion

$$M(n, k) = M(n-1, k) + M(n, k-1) \quad (n > 0, k > 0).$$

One can now prove that for all  $n \geq 0$  and all  $k \geq 0$ ,

$$M(n, k) = \binom{k+n-1}{k, n-1} = \frac{(k+n-1)!}{k!(n-1)!}.$$

The proof is by induction on  $n$ , and is similar to the corresponding proof for  $C(n, k)$ . We leave this proof as an exercise.

If desired, we can use the recursion to compute values of  $M(n, k)$ . Here we use a left-justified table of entries in which the  $n$ th row contains the numbers  $M(n, 0), M(n, 1), \dots$ . The values in the top row (where  $n = 0$ ) and in the left column (where  $k = 0$ ) are given by the initial conditions. Each remaining entry in the table is the sum of the number directly above it and the number directly to its left. See Figure 2.5. The reader will perceive that this is merely a shifted version of Pascal's Triangle.

**2.27. Recursion for Multinomial Coefficients.** Let  $n_1, \dots, n_s$  be nonnegative integers that add to  $n$ . Let  $\{a_1, \dots, a_s\}$  be a given  $s$ -letter alphabet, and let  $C(n; n_1, \dots, n_s) = |\mathcal{R}(a_1^{n_1} \cdots a_s^{n_s})|$  be the number of  $n$ -letter words that are rearrangements of  $n_i$  copies of  $a_i$  for  $1 \leq i \leq s$ . We proved in §1.9 that

$$C(n; n_1, \dots, n_s) = \binom{n}{n_1, \dots, n_s} = \frac{n!}{n_1! n_2! \cdots n_s!}.$$

We now give a new proof of this result using recursions.

Assume first that every  $n_i$  is positive. For  $1 \leq i \leq s$ , let  $T_i$  be the set of words in  $T = \mathcal{R}(a_1^{n_1} \cdots a_s^{n_s})$  that begin with the letter  $a_i$ .  $T$  is the disjoint union of the sets  $T_i$ . To build a typical word  $w \in T_i$ , we start with the letter  $a_i$  and then append any element of

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
$n = 0 :$	1	0	0	0	0	0	0
$n = 1 :$	1	1	1	1	1	1	1
$n = 2 :$	1	2	3	4	5	6	7
$n = 3 :$	1	3	6	10	15	21	28
$n = 4 :$	1	4	10	20	35	56	84
$n = 5 :$	1	5	15	35	70	126	210
$n = 6 :$	1	6	21	56	126	252	462

**FIGURE 2.5**

Table for computing  $M(n, k)$ .

$\mathcal{R}(a_1^{n_1} \cdots a_i^{n_i-1} \cdots a_s^{n_s})$ . There are  $C(n-1; n_1, \dots, n_i-1, \dots, n_s)$  ways to do this. Hence, by the sum rule,

$$C(n; n_1, \dots, n_s) = \sum_{i=1}^s C(n-1; n_1, \dots, n_i-1, \dots, n_s).$$

If we adopt the convention that  $C(n; n_1, \dots, n_s) = 0$  whenever any  $n_i$  is negative, then this recursion holds (with the same proof) for all choices of  $n_i \geq 0$  and  $n > 0$ . The initial condition is

$$C(0; 0, 0, \dots, 0) = 1,$$

since the empty word is the unique rearrangement of zero copies of the given letters.

Now let us prove that

$$C(n; n_1, \dots, n_s) = \frac{n!}{\prod_{k=1}^s n_k!}$$

by induction on  $n$ . In the base case,  $n = n_1 = \cdots = n_s = 0$ , and the desired formula follows from the initial condition. For the induction step, assume that  $n > 0$  and that

$$C(n-1; m_1, \dots, m_s) = \frac{(n-1)!}{\prod_{k=1}^s m_k!}$$

whenever  $m_1 + \cdots + m_s = n-1$ . Assume that we are given integers  $n_k \geq 0$  that sum to  $n$ . Now, using the recursion and induction hypothesis, we compute as follows:

$$\begin{aligned}
C(n; n_1, \dots, n_s) &= \sum_{k=1}^s C(n-1; n_1, \dots, n_k-1, \dots, n_s) \\
&= \sum_{k=1}^s \chi(n_k > 0) \frac{(n-1)!}{(n_k-1)! \prod_{j \neq k} n_j!} \\
&= \sum_{k=1}^s \chi(n_k > 0) \frac{(n-1)! n_k}{\prod_{j=1}^s n_j!} = \sum_{k=1}^s \frac{(n-1)! n_k}{\prod_{j=1}^s n_j!} \\
&= \frac{(n-1)!}{\prod_{j=1}^s n_j!} \left[ \sum_{k=1}^s n_k \right] = \frac{n!}{\prod_{j=1}^s n_j!}.
\end{aligned}$$

## 2.6 Recursions for Lattice Paths

Recursive techniques allow us to count many collections of lattice paths. We first consider the situation of lattice paths in a rectangle.

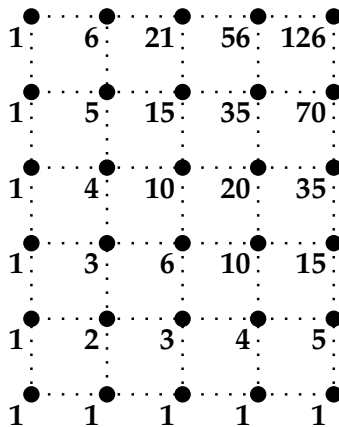
**2.28. Recursion for Paths in a Rectangle.** For  $a, b \geq 0$ , let  $L(a, b)$  be the number of lattice paths from the origin to  $(a, b)$ . We have  $L(a, 0) = L(0, b) = 1$  for all  $a, b \geq 0$ . If  $a > 0$  and  $b > 0$ , note that any lattice path ending at  $(a, b)$  arrives there via an east step or a north step. We obtain lattice paths of the first kind by taking any lattice path ending at  $(a - 1, b)$  and appending an east step. We obtain lattice paths of the second kind by taking any lattice path ending at  $(a, b - 1)$  and appending a north step. Hence, by the sum rule,

$$L(a, b) = L(a - 1, b) + L(a, b - 1) \quad (a, b > 0).$$

One can now show (by induction on  $a + b$ ) that

$$L(a, b) = \binom{a+b}{a, b} = \frac{(a+b)!}{a!b!} \quad (a, b \geq 0).$$

We can visually display and calculate the numbers  $L(a, b)$  by labeling each lattice point  $(a, b)$  with the number  $L(a, b)$ . The initial conditions say that the lattice points on the axes are labeled 1. The recursion says that the label of some point  $(a, b)$  is the sum of the labels of the point  $(a - 1, b)$  to its immediate left and the point  $(a, b - 1)$  immediately below it. See Figure 2.6.



**FIGURE 2.6**

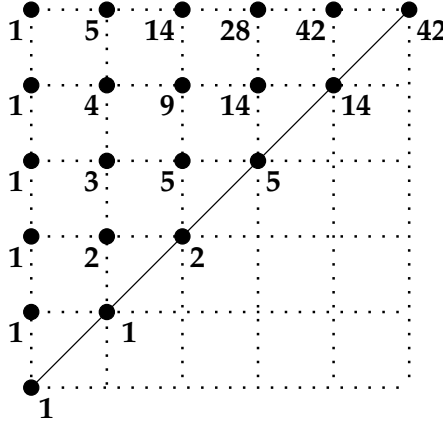
Recursive enumeration of lattice paths.

By modifying the boundary conditions, we can adapt the recursion in the previous example to count more complicated collections of lattice paths.

**2.29. Recursion for Paths in a Triangle.** For  $b \geq a \geq 0$ , let  $T(a, b)$  be the number of lattice paths from the origin to  $(a, b)$  that always stay weakly above the line  $y = x$ . (In particular,  $T(n, n)$  is the number of Dyck paths of order  $n$ .) By the same argument used above, we have

$$T(a, b) = T(a - 1, b) + T(a, b - 1) \quad (b > a > 0).$$



**FIGURE 2.7**

Recursive enumeration of lattice paths in a triangle.

On the other hand, when  $a = b > 0$ , a lattice path can only reach  $(a, b) = (a, a)$  by taking an east step, since the point  $(a, b - 1)$  lies below  $y = x$ . Thus,

$$T(a, a) = T(a - 1, a) \quad (a > 0).$$

The initial conditions are  $T(0, b) = 1$  for all  $b \geq 0$ . Figure 2.7 shows how to compute the numbers  $T(a, b)$  by drawing a picture.

It turns out that there is an explicit closed formula for the numbers  $T(a, b)$ .

**2.30. Theorem: Ballot Numbers.** For  $b \geq a \geq 0$ , the number of lattice paths from the origin to  $(a, b)$  that always stay weakly above the line  $y = x$  is

$$\frac{b - a + 1}{b + a + 1} \binom{a + b + 1}{a}.$$

In particular, the number of Dyck paths of order  $n$  is

$$\frac{1}{2n + 1} \binom{2n + 1}{n} = C_n.$$

*Proof.* We show that

$$T(a, b) = \frac{b - a + 1}{b + a + 1} \binom{a + b + 1}{a}$$

by induction on  $a + b$ . If  $a + b = 0$ , so that  $a = b = 0$ , then  $T(0, 0) = 1 = \frac{0-0+1}{0+0+1} \binom{0+0+1}{0}$ . Now assume that  $a + b > 0$  and that  $T(c, d) = \frac{d-c+1}{d+c+1} \binom{c+d+1}{c}$  whenever  $d \geq c \geq 0$  and  $c + d < a + b$ . To prove the desired formula for  $T(a, b)$ , we consider cases based on the recursions and initial conditions. First, if  $a = 0$  and  $b \geq 0$ , we have  $T(a, b) = 1 = \frac{b-0+1}{b+0+1} \binom{0+b+1}{0}$ . Second, if  $a = b > 0$ , we have

$$\begin{aligned} T(a, b) &= T(a, a) = T(a - 1, a) = \frac{2}{2a} \binom{2a}{a - 1} \\ &= \frac{(2a)!}{a!(a + 1)!} = \frac{1}{2a + 1} \binom{2a + 1}{a} \\ &= \frac{b - a + 1}{b + a + 1} \binom{a + b + 1}{a}. \end{aligned}$$

Third, if  $b > a > 0$ , we have

$$\begin{aligned}
 T(a, b) &= T(a-1, b) + T(a, b-1) = \frac{b-a+2}{a+b} \binom{a+b}{a-1} + \frac{b-a}{a+b} \binom{a+b}{a} \\
 &= \frac{(b-a+2)(a+b-1)!}{(a-1)!(b+1)!} + \frac{(b-a)(a+b-1)!}{a!b!} \\
 &= \left[ \frac{a(b-a+2)}{a+b} + \frac{(b-a)(b+1)}{a+b} \right] \frac{(a+b)!}{a!(b+1)!} \\
 &= \left[ \frac{ab-a^2+2a+b^2-ab+b-a}{a+b} \right] \frac{(a+b+1)!}{(b+a+1)a!(b+1)!} \\
 &= \left[ \frac{(b-a+1)(a+b)}{a+b} \right] \frac{1}{b+a+1} \binom{a+b+1}{a} \\
 &= \frac{b-a+1}{b+a+1} \binom{a+b+1}{a}. \quad \square
 \end{aligned}$$

The numbers  $T(a, b)$  in the previous theorem are called *ballot numbers*, for the following reason. Let  $\pi \in \{N, E\}^{a+b}$  be a lattice path counted by  $T(a, b)$ . Imagine that  $a+b$  people are voting for two candidates (“candidate N” and “candidate E”) by casting an ordered sequence of  $a+b$  ballots. The path  $\pi$  records this sequence of ballots as follows:  $\pi_j = N$  if the  $j$ th person votes for candidate N, and  $\pi_j = E$  if the  $j$ th person votes for candidate E. The condition that  $\pi$  stays weakly above  $y = x$  means that candidate N always has at least as many votes as candidate E at each stage in the election process. The condition that  $\pi$  ends at  $(a, b)$  means that candidate N has  $b$  votes and candidate E has  $a$  votes at the end of the election.

Returning to lattice paths, suppose we replace the boundary line  $y = x$  by the line  $y = mx$  (where  $m$  is any positive integer). We can then derive the following more general result.

**2.31. Theorem:  $m$ -Ballot Numbers.** Let  $m$  be a fixed positive integer. For  $b \geq ma \geq 0$ , the number of lattice paths from the origin to  $(a, b)$  that always stay weakly above the line  $y = mx$  is

$$\frac{b-ma+1}{b+a+1} \binom{a+b+1}{a}.$$

In particular, the number of such paths ending at  $(n, mn)$  is

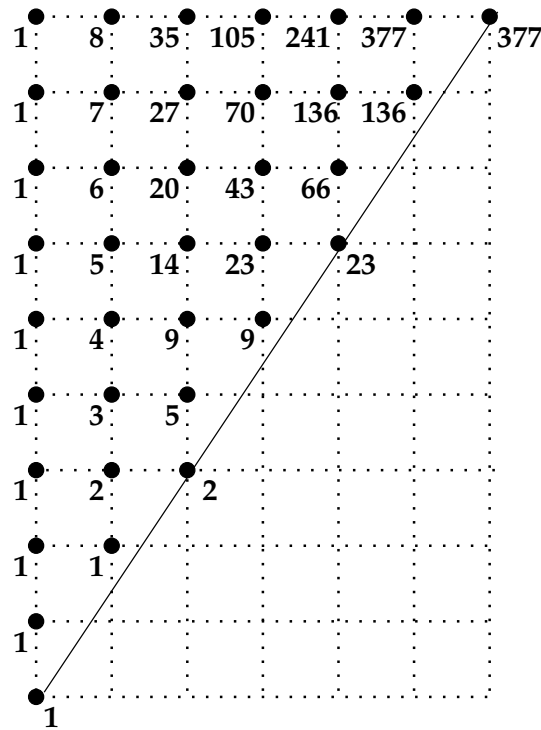
$$\frac{1}{(m+1)n+1} \binom{(m+1)n+1}{n}.$$

*Proof.* Let  $T_m(a, b)$  be the number of paths ending at  $(a, b)$  that never go below  $y = mx$ . Arguing as before, we have  $T_m(0, b) = 1$  for all  $b \geq 0$ ;  $T_m(a, b) = T_m(a-1, b) + T_m(a, b-1)$  whenever  $b > ma > 0$ ; and  $T_m(a, ma) = T_m(a-1, ma)$  since the point  $(a, ma-1)$  lies below the line  $y = mx$ . One now proves that

$$T_m(a, b) = \frac{b-ma+1}{b+a+1} \binom{a+b+1}{a}$$

by induction on  $a+b$ . The proof is similar to the one given above, so we leave it as an exercise. For a bijective proof of this theorem, see 12.92.  $\square$

When the slope  $m$  of the boundary line  $y = mx$  is not an integer, we cannot use the formula in the preceding theorem. Nevertheless, the recursion (with appropriate initial



conditions) can still be used to count lattice paths bounded below by this line. For example, Figure 2.8 illustrates the enumeration of lattice paths from  $(0, 0)$  to  $(6, 9)$  that always stay weakly above  $y = (3/2)x$ .

We end this section with a general recursion for counting lattice paths in a given region.

**2.32. Theorem: General Lattice Path Recursion.** Suppose  $V$  is a given set of lattice points in  $\mathbb{N} \times \mathbb{N}$  containing the origin. For  $(a, b) \in V$ , let  $T_V(a, b)$  be the number of lattice paths from the origin to  $(a, b)$  that visit only lattice points in  $V$ . Then  $T_V(0, 0) = 1$  and

$$T_V(a, b) = T_V(a-1, b)\chi((a-1, b) \in V) + T_V(a, b-1)\chi((a, b-1) \in V) \quad \text{for } (a, b) \neq (0, 0).$$

The proof is immediate from the sum rule. Figure 2.9 illustrates the use of this recursion to count lattice paths contained in an irregular region. In the figure, lattice points in  $V$  are drawn as closed circles, while X's indicate certain forbidden lattice points that the path is not allowed to use.

## 2.7 Catalan Recursions

The recursions from the previous section provide one way of computing Catalan numbers, which are a special case of ballot numbers. This section discusses another recursion that involves only the Catalan numbers. This “convolution recursion” comes up in many settings, thus leading to many different combinatorial interpretations for the Catalan numbers.

**2.33. Theorem: Catalan Recursion.** The Catalan numbers  $C_n = \frac{1}{n+1} \binom{2n}{n}$  satisfy the recursion

$$C_n = \sum_{k=1}^n C_{k-1} C_{n-k} \quad (n > 0)$$

and initial condition  $C_0 = 1$ .

*Proof.* Recall from 1.56 that  $C_n$  is the number of Dyck paths of order  $n$ . There is one Dyck path of order 0, so  $C_0 = 1$ . Fix  $n > 0$ , and let  $A$  be the set of Dyck paths ending at  $(n, n)$ . For  $1 \leq k \leq n$ , let

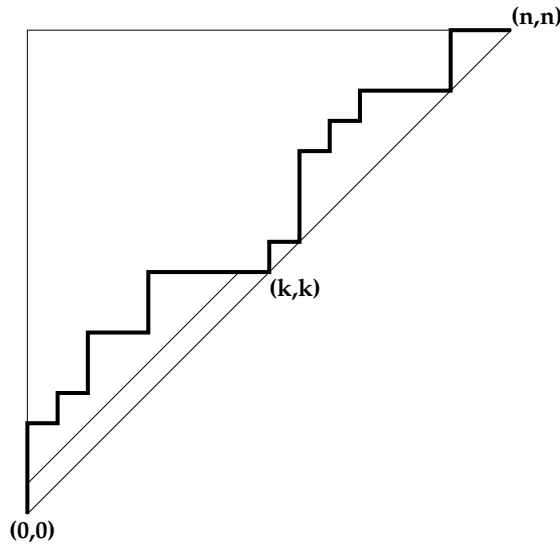
$$A_k = \{\pi \in A : (k, k) \in \pi \text{ and } (j, j) \notin \pi \text{ for } 0 < j < k\}.$$

In other words,  $A_k$  consists of the Dyck paths of order  $n$  that return to the diagonal line  $y = x$  for the first time at the point  $(k, k)$ . See Figure 2.10. Suppose  $w$  is the word in  $\{N, E\}^{2n}$  that encodes a path  $\pi \in A_k$ . Inspection of Figure 2.10 shows that we have the factorization  $w = Nw_1Ew_2$ , where  $w_1$  encodes a Dyck path of order  $k-1$  (starting at  $(0, 1)$  in the figure) and  $w_2$  encodes a Dyck path of order  $n-k$  (starting at  $(k, k)$  in the figure). We can uniquely construct all paths in  $A_k$  by choosing  $w_1$  and  $w_2$  and then setting  $w = Nw_1Ew_2$ . There are  $C_{k-1}$  choices for  $w_1$  and  $C_{n-k}$  choices for  $w_2$ . By the product rule and sum rule,

$$C_n = |A| = \sum_{k=1}^n |A_k| = \sum_{k=1}^n C_{k-1} C_{n-k}.$$

□

The next result shows that the Catalan recursion uniquely determines the Catalan numbers.

**FIGURE 2.10**

Proving the Catalan recursion by analyzing the first return to  $y = x$ .

**2.34. Theorem.** Suppose  $(d_n : n \geq 0)$  is a sequence such that  $d_0 = 1$  and

$$d_n = \sum_{k=1}^n d_{k-1} d_{n-k} \quad (n > 0).$$

Then  $d_n = C_n = \frac{1}{n+1} \binom{2n}{n}$  for all  $n \geq 0$ .

*Proof.* We argue by strong induction. For  $n = 0$ , we have  $d_0 = 1 = C_0$ . Assume that  $n > 0$  and that  $d_m = C_m$  for all  $m < n$ . Then

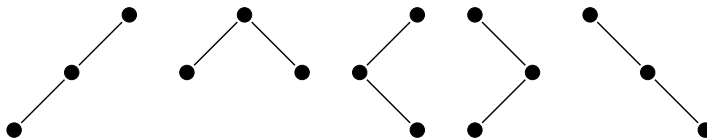
$$d_n = \sum_{k=1}^n d_{k-1} d_{n-k} = \sum_{k=1}^n C_{k-1} C_{n-k} = C_n. \quad \square$$

We can now prove that various collections of objects are counted by the Catalan numbers. One proof method sets up a bijection between such objects and other objects (like Dyck paths) that are already known to be counted by Catalan numbers. A second proof method shows that the new collections of objects satisfy the Catalan recursion. We illustrate both methods in the examples below.

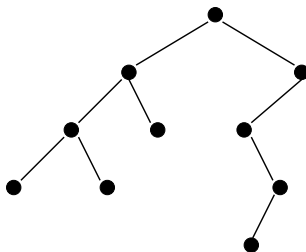
**2.35. Example: Balanced Parentheses.** For  $n \geq 0$ , let  $BP_n$  be the set of all words consisting of  $n$  left parentheses and  $n$  right parentheses, such that every left parenthesis can be matched with a right parenthesis later in the word. For example,  $BP_3$  consists of the following five words:

$$((())), \quad (())(), \quad ()(()), \quad (()()), \quad ()()().$$

We show that  $|BP_n| = C_n$  for all  $n$  by exhibiting a bijection between  $BP_n$  and the set of Dyck paths of order  $n$ . Given  $w \in BP_3$ , replace each left parenthesis by N (which encodes a north step) and each right parenthesis by E (which encodes an east step). One can check that a string  $w$  of  $n$  left and  $n$  right parentheses is balanced iff for every  $i \leq 2n$ , the number

**FIGURE 2.11**

The five binary trees with three nodes.

**FIGURE 2.12**

A binary tree with ten nodes.

of left parentheses in the prefix  $w_1 w_2 \cdots w_i$  weakly exceeds the number of right parentheses in this prefix. Converting to north and east steps, this condition means that no lattice point on the path lies strictly below the line  $y = x$ . Thus we have mapped each  $w \in BP_n$  to a Dyck path. This map is a bijection, so  $|BP_n| = C_n$ .

**2.36. Example: Binary Trees.** We recursively define the set of *binary trees with  $n$  nodes* as follows. The empty set is the unique binary tree with 0 nodes. If  $T_1$  is a binary tree with  $h$  nodes and  $T_2$  is a binary tree with  $k$  nodes, then the ordered triple  $T = (\bullet, T_1, T_2)$  is a binary tree with  $h + k + 1$  nodes. By definition, all binary trees arise by a finite number of applications of these rules. If  $T = (\bullet, T_1, T_2)$  is a binary tree, we call  $T_1$  the *left subtree* of  $T$  and  $T_2$  the *right subtree* of  $T$ . Note that  $T_1$  or  $T_2$  (or both) may be empty. We can draw a picture of a nonempty binary tree  $T$  as follows. First, draw a *root node* of the binary tree at the top of the picture. If  $T_1$  is nonempty, draw an edge leading down and left from the root node, and then draw the picture of  $T_1$ . If  $T_2$  is nonempty, draw an edge leading down and right from the root node, and then draw the picture of  $T_2$ . For example, Figure 2.11 displays the five binary trees with three nodes. Figure 2.12 depicts a larger binary tree that is formally represented by the sequence

$$T = (\bullet, (\bullet, (\bullet, (\bullet, \emptyset, \emptyset), (\bullet, \emptyset, \emptyset)), (\bullet, \emptyset, \emptyset)), (\bullet, (\bullet, \emptyset, (\bullet, (\bullet, \emptyset, \emptyset), \emptyset)), \emptyset)).$$

Let  $BT_n$  denote the set of binary trees with  $n$  nodes. We show that  $|BT_n| = C_n$  for all  $n$  by verifying that the sequence  $(|BT_n| : n \geq 0)$  satisfies the Catalan recursion. First,  $|BT_0| = 1$  by definition. Second, suppose  $n \geq 1$ . By the recursive definition of binary trees, we can uniquely construct a typical element of  $BT_n$  as follows. Fix  $k$  with  $1 \leq k \leq n$ . Choose a tree  $T_1 \in BT_{k-1}$  with  $k-1$  nodes. Then choose a tree  $T_2 \in BT_{n-k}$  with  $n-k$  nodes. We assemble these trees (together with a new root node) to get a binary tree  $T = (\bullet, T_1, T_2)$  with  $(k-1) + 1 + (n-k) = n$  nodes. By the sum and product rules, we have

$$|BT_n| = \sum_{k=1}^n |BT_{k-1}| |BT_{n-k}|.$$

It follows from 2.34 that  $|BT_n| = C_n$  for all  $n \geq 0$ .

**2.37. Example: 231-avoiding permutations.** Suppose  $w = w_1w_2 \cdots w_n$  is a permutation of  $n$  distinct integers. We say that  $w$  is *231-avoiding* iff there do not exist indices  $i < k < p$  such that  $w_p < w_i < w_k$ . This means that no three-element subsequence  $w_i \dots w_k \dots w_p$  in  $w$  has the property that  $w_p$  is the smallest number in  $\{w_i, w_k, w_p\}$  and  $w_k$  is the largest number in  $\{w_i, w_k, w_p\}$ . For example, when  $n = 4$ , there are fourteen 231-avoiding permutations of  $\{1, 2, 3, 4\}$ :

1234, 1243, 1324, 1423, 1432, 2134, 2143,

3124, 3214, 4123, 4132, 4213, 4312, 4321.

The following ten permutations do contain occurrences of the pattern 231:

2314, 2341, 2431, 4231, 3421, 3412, 3142, 3241, 2413.

Let  $S_n^{231}$  be the set of 231-avoiding permutations of  $\{1, 2, \dots, n\}$ . We prove that  $|S_n^{231}| = C_n$  for all  $n \geq 0$  by verifying the Catalan recursion. First,  $|S_0^{231}| = 1 = C_0$  since the empty permutation is certainly 231-avoiding. Next, suppose  $n > 0$ . We construct a typical object  $w \in S_n^{231}$  as follows. Consider cases based on the position of the letter  $n$  in  $w$ . Say  $w_k = n$ . For all  $i < k$  and all  $p > k$ , we must have  $w_i < w_p$ ; otherwise, the subsequence  $w_i, w_k = n, w_p$  would be an occurrence of the forbidden 231 pattern. Assuming that  $w_i < w_p$  whenever  $i < k < p$ , one checks that  $w = w_1w_2 \cdots w_n$  is 231-avoiding iff  $w_1w_2 \cdots w_{k-1}$  is 231-avoiding and  $w_{k+1} \cdots w_n$  is 231-avoiding. Thus, for a fixed  $k$ , we can construct  $w$  by choosing an arbitrary 231-avoiding permutation  $w'$  of the  $k-1$  letters  $\{1, 2, \dots, k-1\}$  in  $|S_{k-1}^{231}|$  ways, then choosing an arbitrary 231-avoiding permutation  $w''$  of the  $n-k$  letters  $\{k, \dots, n-1\}$  in  $|S_{n-k}^{231}|$  ways, and finally letting  $w$  be the concatenation of  $w'$ , the letter  $n$ , and  $w''$ . By the sum and product rules, we have

$$|S_n^{231}| = \sum_{k=1}^n |S_{k-1}^{231}| |S_{n-k}^{231}|.$$

By 2.34,  $|S_n^{231}| = C_n$  for all  $n \geq 0$ .

**2.38. Example:  $\tau$ -avoiding permutations.** Let  $\tau : \{1, 2, \dots, k\} \rightarrow \{1, 2, \dots, k\}$  be a fixed permutation of  $k$  letters. A permutation  $w$  of  $\{1, 2, \dots, n\}$  is called  *$\tau$ -avoiding* iff there do not exist indices  $1 \leq i(1) < i(2) < \cdots < i(k) \leq n$  such that

$$w_{i(\tau^{-1}(1))} < w_{i(\tau^{-1}(2))} < \cdots < w_{i(\tau^{-1}(k))}.$$

This means that no subsequence of  $k$  entries of  $w$  consists of numbers in the same relative order as the numbers  $\tau_1, \tau_2, \dots, \tau_k$ . For instance,  $w = 15362784$  is not 2341-avoiding, since the subsequence 5684 matches the pattern 2341 (as does the subsequence 5674). On the other hand,  $w$  is 4321-avoiding, since there is no descending subsequence of  $w$  of length 4. Let  $S_n^\tau$  denote the set of  $\tau$ -avoiding permutations of  $\{1, 2, \dots, n\}$ .

For general  $\tau$ , the enumeration of  $\tau$ -avoiding permutations is an extremely difficult problem that has stimulated much research in recent years. On the other hand, if  $\tau$  is a permutation of  $k = 3$  letters, then the number of  $\tau$ -avoiding permutations of length  $n$  is always the Catalan number  $C_n$ , for all six possible choices of  $\tau$ . We have already proved this in the last example for  $\tau = 231$ . The arguments in that example readily adapt to prove the Catalan recursion for  $\tau = 132$ ,  $\tau = 213$ , and  $\tau = 312$ . However, more subtle arguments are needed to prove this result for  $\tau = 123$  and  $\tau = 321$  (see 12.65).

**2.39. Remark.** Let  $(A_n : n \geq 0)$  and  $(B_n : n \geq 0)$  be two families of combinatorial objects such that  $|A_n| = C_n = |B_n|$  for all  $n$ . Suppose that we have an explicit bijective proof that the numbers  $|A_n|$  satisfy the Catalan recursion. This means that we can describe a bijection  $g_n$  between the set  $A_n$  and the disjoint union of the sets  $A_{k-1} \times A_{n-k}$  for  $k = 1, 2, \dots, n$ . (Such a bijection is usually implicit in an argument involving the sum and product rules.) Suppose we have similar bijections  $h_n$  for the sets  $B_n$ . We can combine these bijections to obtain recursively defined bijections  $f_n : A_n \rightarrow B_n$ . First, there is a unique bijection  $f_0 : A_0 \rightarrow B_0$ , since  $|A_0| = 1 = |B_0|$ . Second, assume that  $f_m : A_m \rightarrow B_m$  has already been defined for all  $m < n$ . Define  $f_n : A_n \rightarrow B_n$  as follows. Given  $x \in A_n$ , suppose  $g_n(x) = (k, y, z)$  where  $1 \leq k \leq n$ ,  $y \in A_{k-1}$ , and  $z \in A_{n-k}$ . Set

$$f_n(x) = h_n^{-1}((k, f_{k-1}(y), f_{n-k}(z))).$$

The inverse map is defined analogously.

For example, let us recursively define a bijection  $\phi$  from the set of binary trees to the set of Dyck paths such that trees with  $n$  nodes map to paths of order  $n$ . Linking together the first-return recursion for Dyck paths with the left/right-subtree recursion for binary trees as discussed in the previous paragraph, we obtain the rule

$$\phi(\emptyset) = \text{the empty word (denoted } \epsilon); \quad \phi((\bullet, T_1, T_2)) = N\phi(T_1)E\phi(T_2).$$

For example, the one-node tree  $(\bullet, \emptyset, \emptyset)$  maps to the Dyck path  $N\epsilon E\epsilon = NE$ . It then follows that

$$\phi(\bullet, (\bullet, \emptyset, \emptyset), \emptyset) = N(NE)E\epsilon = NNEE;$$

$$\phi(\bullet, \emptyset, (\bullet, \emptyset, \emptyset)) = N\epsilon E(NE) = NENE;$$

$$\phi(\bullet, (\bullet, \emptyset, \emptyset), (\bullet, \emptyset, \emptyset)) = N(NE)E(NE) = NNEENE;$$

and so on. Figure 2.13 illustrates the recursive computation of  $\phi(T)$  for the binary tree  $T$  shown in Figure 2.12.

As another example, let us recursively define a bijection  $\psi$  from the set of binary trees to the set of 231-avoiding permutations such that trees with  $n$  nodes map to permutations of  $n$  letters. Linking together the two proofs of the Catalan recursion for binary trees and 231-avoiding permutations, we obtain the rule

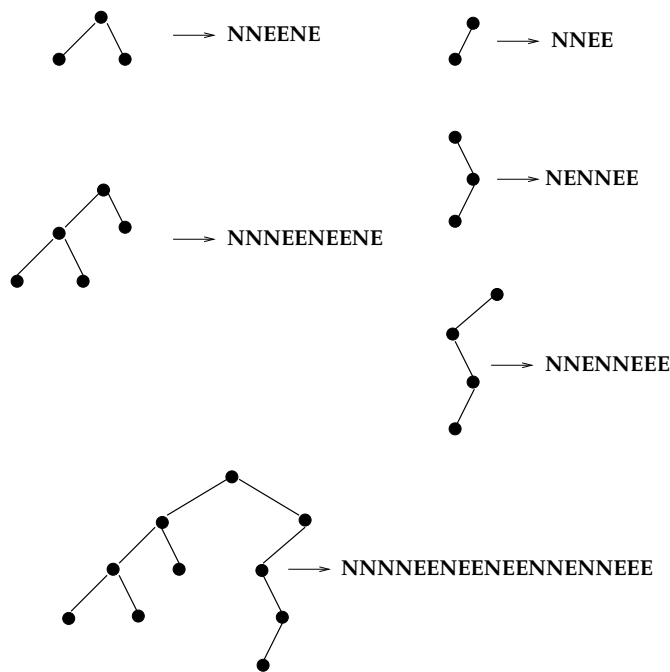
$$\psi(\emptyset) = \epsilon, \quad \psi((\bullet, T_1, T_2)) = \psi(T_1)n\psi'(T_2),$$

where  $\psi'(T_2)$  is the permutation obtained by increasing each entry of  $\psi(T_2)$  by  $k-1 = |T_1|$ . Figure 2.14 illustrates the recursive computation of  $\psi(T)$  for the binary tree  $T$  shown in Figure 2.12.

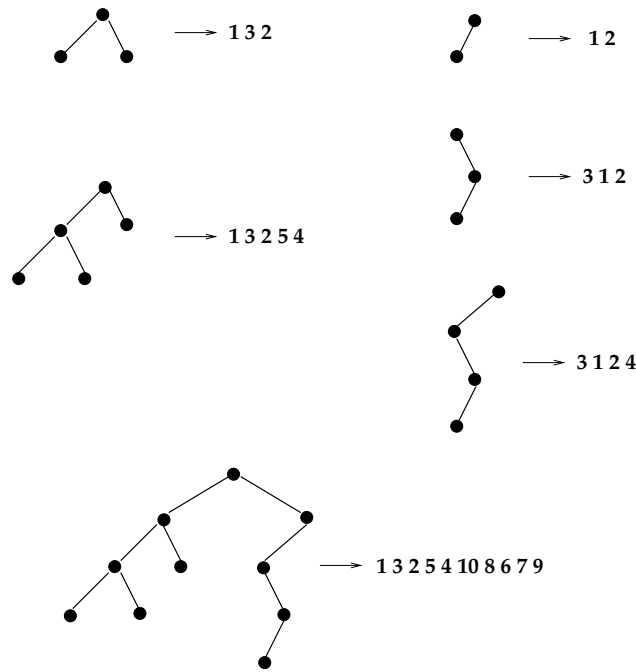
## 2.8 Integer Partitions

**2.40. Definition: Integer Partitions.** Let  $n$  be a nonnegative integer. An *integer partition* of  $n$  is a sequence  $\mu = (\mu_1, \mu_2, \dots, \mu_k)$  of positive integers such that  $\mu_1 + \mu_2 + \dots + \mu_k = n$  and  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_k$ . Each  $\mu_i$  is called a *part* of the partition. Let  $p(n)$  be the number of integer partitions of  $n$ , and let  $p(n, k)$  be the number of integer partitions of  $n$  into exactly  $k$  parts. If  $\mu$  is a partition of  $n$  into  $k$  parts, we write  $|\mu| = n$  and  $\ell(\mu) = k$  and say that  $\mu$  has *area*  $n$  and *length*  $k$ . Let  $\text{Par}$  denote the set of all integer partitions.





**FIGURE 2.13**  
Mapping binary trees to Dyck paths.



**FIGURE 2.14**  
Mapping binary trees to 231-avoiding permutations.

**2.41. Example.** The integer partitions of 5 are

$$(5), (4, 1), (3, 2), (3, 1, 1), (2, 2, 1), (2, 1, 1, 1), (1, 1, 1, 1, 1).$$

Thus,  $p(5) = 7$ ,  $p(5, 1) = 1$ ,  $p(5, 2) = 2$ ,  $p(5, 3) = 2$ ,  $p(5, 4) = 1$ , and  $p(5, 5) = 1$ . As another example, the empty sequence is the unique integer partition of 0, so  $p(0) = 1 = p(0, 0)$ .

An integer partition of  $n$  is a composition of  $n$  in which the parts appear in weakly decreasing order. Informally, we can think of an integer partition of  $n$  as a way of writing  $n$  as a sum of positive integers where the order of the summands does not matter.

We know from 1.41 that there are  $2^{n-1}$  compositions of  $n$ . One might hope for a similar explicit formula for  $p(n)$ . Such a formula does exist (see 2.49 below), but it is extraordinarily complicated. Fortunately, the numbers  $p(n, k)$  do satisfy a nice recursion.

**2.42. Theorem: Recursion for Integer Partitions.** Let  $p(n, k)$  be the number of integer partitions of  $n$  into  $k$  parts. Then

$$p(n, k) = p(n-1, k-1) + p(n-k, k) \quad (n, k > 0).$$

The initial conditions are  $p(n, k) = 0$  for  $k > n$  or  $k < 0$ ,  $p(n, 0) = 0$  for  $n > 0$ , and  $p(0, 0) = 1$ .

*Proof.* For all  $i, j$ , let  $P(i, j)$  be the set of integer partitions of  $i$  into  $j$  parts. We have  $|P(i, j)| = p(i, j)$ . Fix  $n, k > 0$ . The set  $P(n, k)$  is the disjoint union of the two sets

$$\begin{aligned} Q &= \{(\mu_1, \dots, \mu_k) \in P(n, k) : \mu_k = 1\}, \\ R &= \{(\mu_1, \dots, \mu_k) \in P(n, k) : \mu_k > 1\}. \end{aligned}$$

On one hand, the map  $(\mu_1, \dots, \mu_k) \mapsto (\mu_1, \dots, \mu_{k-1})$  is a bijection from  $Q$  onto  $P(n-1, k-1)$  with inverse  $(\nu_1, \dots, \nu_{k-1}) \mapsto (\nu_1, \dots, \nu_{k-1}, 1)$ . So  $|Q| = |P(n-1, k-1)| = p(n-1, k-1)$ . On the other hand, the map  $(\mu_1, \dots, \mu_k) \mapsto (\mu_1-1, \mu_2-1, \dots, \mu_k-1)$  is a bijection from  $R$  onto  $P(n-k, k)$  with inverse  $(\rho_1, \dots, \rho_k) \mapsto (\rho_1+1, \dots, \rho_k+1)$ . So  $|R| = |P(n-k, k)| = p(n-k, k)$ . The sum rule now gives

$$p(n, k) = |P(n, k)| = |Q| + |R| = p(n-1, k-1) + p(n-k, k). \quad \square$$

**2.43. Theorem: Dual Recursion for Partitions.** Let  $p'(n, k)$  be the number of integer partitions of  $n$  with first part  $k$ . Then

$$p'(n, k) = p'(n-1, k-1) + p'(n-k, k) \quad (n, k > 0).$$

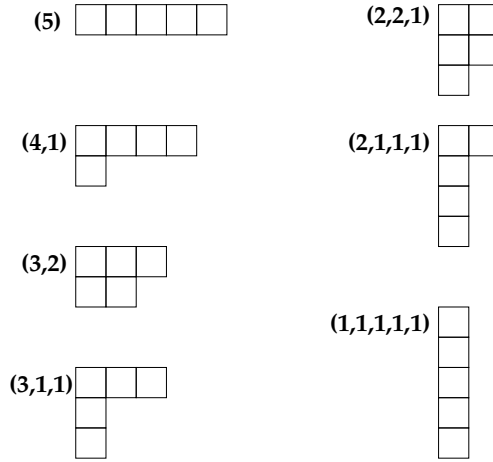
The initial conditions are  $p'(n, k) = 0$  for  $k > n$  or  $k < 0$ ,  $p'(n, 0) = 0$  for  $n > 0$ , and (by convention)  $p'(0, 0) = 1$ .

*Proof.* For all  $i, j$ , let  $P'(i, j)$  be the set of integer partitions of  $i$  with first part  $j$ . We have  $|P'(i, j)| = p'(i, j)$ . Fix  $n, k > 0$ . The set  $P'(n, k)$  is the disjoint union of the two sets

$$\begin{aligned} Q &= \{(\mu_1 = k, \mu_2, \dots, \mu_s) \in P'(n, k) : \mu_1 > \mu_2\} \\ R &= \{(\mu_1 = k, \mu_2, \dots, \mu_s) \in P'(n, k) : \mu_1 = \mu_2\}. \end{aligned}$$

(If  $\mu$  has only one part, we take  $\mu_2 = 0$  by convention.) On one hand, the map  $(k, \mu_2, \dots, \mu_s) \mapsto (k-1, \mu_2, \dots, \mu_s)$  is a bijection from  $Q$  onto  $P'(n-1, k-1)$  with inverse  $(k-1, \nu_2, \dots, \nu_s) \mapsto (k, \nu_2, \dots, \nu_s)$ . So  $|Q| = |P'(n-1, k-1)| = p'(n-1, k-1)$ . On the other hand, the map  $(k, \mu_2, \dots, \mu_s) \mapsto (\mu_2, \mu_3, \dots, \mu_s)$  is a bijection from  $R$  onto  $P'(n-k, k)$  with inverse  $(\rho_1, \dots, \rho_s) \mapsto (k, \rho_1, \dots, \rho_s)$ . So  $|R| = |P'(n-k, k)| = p'(n-k, k)$ . The sum rule now gives

$$p'(n, k) = |P'(n, k)| = |Q| + |R| = p'(n-1, k-1) + p'(n-k, k). \quad \square$$



**FIGURE 2.15**  
Partition diagrams.

**2.44. Theorem: First Part vs. Number of Parts.** The number of integer partitions of  $n$  into  $k$  parts equals the number of integer partitions of  $n$  with first part  $k$ .

*Proof.* We prove  $p(n, k) = p'(n, k)$  for all  $n$  and all  $k$  by induction on  $n$ . The case  $n = 0$  is true since the initial conditions show that  $p(0, k) = p'(0, k)$  for all  $k$ . Now assume that  $n > 0$  and that  $p(m, k) = p'(m, k)$  for all  $m < n$  and all  $k$ . If  $k = 0$ , then  $p(n, 0) = p'(n, 0)$  follows from the initial conditions. If  $k > 0$ , compute

$$p(n, k) = p(n-1, k-1) + p(n-k, k) = p'(n-1, k-1) + p'(n-k, k) = p'(n, k). \quad \square$$

We now describe a convenient way of visualizing integer partitions.

**2.45. Definition: Diagram of a Partition.** Let  $\mu = (\mu_1, \mu_2, \dots, \mu_k)$  be an integer partition of  $n$ . The *diagram* of  $\mu$  is the set

$$\text{dg}(\mu) = \{(i, j) \in \mathbb{N} \times \mathbb{N} : 1 \leq i \leq k, 1 \leq j \leq \mu_i\}.$$

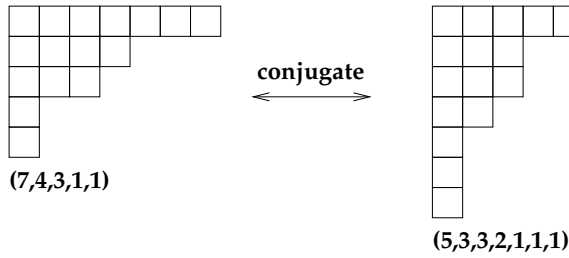
We can make a picture of  $\text{dg}(\mu)$  by drawing an array of  $n$  boxes, with  $\mu_i$  left-justified boxes in row  $i$ . For example, Figure 2.15 illustrates the diagrams for the seven integer partitions of 5. Note that  $|\mu| = \mu_1 + \dots + \mu_k = |\text{dg}(\mu)|$  is the total number of boxes in the diagram of  $\mu$ .

**2.46. Definition: Conjugate Partitions.** Suppose  $\mu$  is an integer partition of  $n$ . The *conjugate partition* of  $\mu$  is the unique integer partition  $\mu'$  of  $n$  satisfying

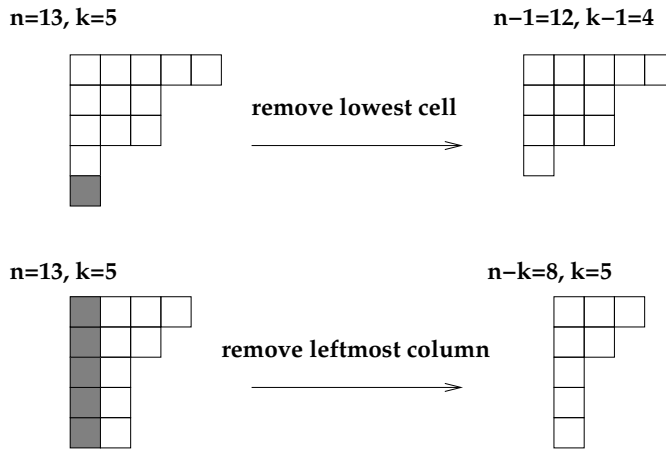
$$\text{dg}(\mu') = \{(j, i) : (i, j) \in \text{dg}(\mu)\}.$$

In other words, we obtain the diagram for  $\mu'$  by interchanging the rows and columns in the diagram for  $\mu$ . For example, Figure 2.16 shows that the conjugate of  $\mu = (7, 4, 3, 1, 1)$  is  $\mu' = (5, 3, 3, 2, 1, 1, 1)$ .

We can now give pictorial proofs of some of the preceding results concerning  $p(n, k)$  and  $p'(n, k)$ . Note that the length of a partition  $\mu$  is the number of rows in  $\text{dg}(\mu)$ , while the first part of  $\mu$  is the number of columns of  $\text{dg}(\mu)$ . Hence, conjugation gives a bijection

**FIGURE 2.16**

Conjugate of a partition.

**FIGURE 2.17**Pictorial proof of the recursion for  $p(n, k)$ .

between the partitions counted by  $p(n, k)$  and the partitions counted by  $p'(n, k)$ , so that  $p(n, k) = p'(n, k)$ . Similarly, consider the proof of the recursion

$$p(n, k) = p(n-1, k-1) + p(n-k, k).$$

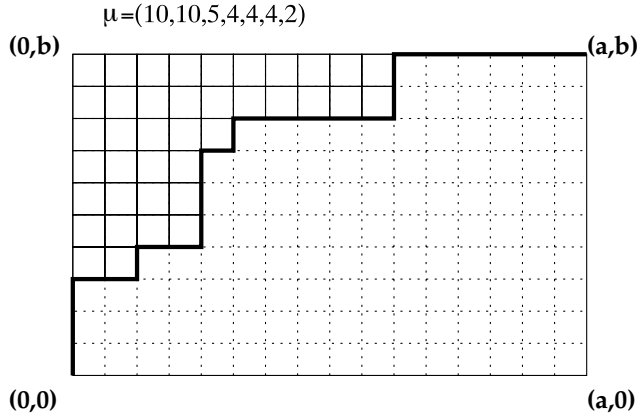
Suppose  $\mu$  is a partition of  $n$  into  $k$  parts. If  $\text{dg}(\mu)$  has one box in the lowest row, we remove this box to get a typical partition counted by  $p(n-1, k-1)$ . If  $\text{dg}(\mu)$  has more than one box in the lowest row, we remove the entire first column of the diagram to get a typical partition counted by  $p(n-k, k)$ . See Figure 2.17.

Our next result counts integer partitions whose diagrams fit in a box with  $b$  rows and  $a$  columns.

**2.47. Theorem: Enumeration of Partitions in a Box.** The number of integer partitions  $\mu$  such that  $\text{dg}(\mu) \subseteq \{1, 2, \dots, b\} \times \{1, 2, \dots, a\}$  is

$$\binom{a+b}{a, b} = \frac{(a+b)!}{a!b!}.$$

*Proof.* We define a bijection between the set of integer partitions in the theorem statement and the set of all lattice paths from the origin to  $(a, b)$ . We draw our partition diagrams

**FIGURE 2.18**

Counting partitions that fit in an  $a \times b$  box.

in the box with corners  $(0,0)$ ,  $(a,0)$ ,  $(0,b)$ , and  $(a,b)$ , as shown in Figure 2.18. Given a partition  $\mu$  whose diagram fits in this box, the southeast boundary of  $\text{dg}(\mu)$  is a lattice path from the origin to  $(a,b)$ . We call this lattice path the *frontier* of  $\mu$  (which depends on  $a$  and  $b$  as well as  $\mu$ ). For example, if  $a = 16$ ,  $b = 10$ , and  $\mu = (10, 10, 5, 4, 4, 4, 2)$ , we see from Figure 2.18 that the frontier of  $\mu$  is

NNNEENEENNENEEEEENNNEEEEE.

Conversely, given any lattice path ending at  $(a,b)$ , the set of lattice squares northwest of this path in the box uniquely determines the diagram of an integer partition. We already know that the number of lattice paths from the origin to  $(a,b)$  is  $\binom{a+b}{a,b}$ , so the theorem follows.  $\square$

**2.48. Remark: Euler's Partition Recursion.** Our recursion for  $p(n, k)$  gives a quick method for computing the quantities  $p(n, k)$  and  $p(n) = \sum_{k=1}^n p(n, k)$ . One may ask whether the numbers  $p(n)$  satisfy any recursion. In fact, Euler's study of the infinite product  $\prod_{i=1}^{\infty} (1 - x^i)$  leads to the following recursion for  $p(n)$ :

$$\begin{aligned} p(n) &= \sum_{m=1}^{\infty} (-1)^{m-1} [p(n - m(3m-1)/2) + p(n - m(3m+1)/2)] \\ &= p(n-1) + p(n-2) - p(n-5) - p(n-7) + p(n-12) + p(n-15) \\ &\quad - p(n-22) - p(n-26) + p(n-35) + p(n-40) - p(n-51) - p(n-57) + \cdots \end{aligned}$$

The initial conditions are  $p(0) = 1$  and  $p(j) = 0$  for all  $j < 0$ . It follows that, for each fixed  $n$ , the recursive expression for  $p(n)$  is really a finite sum, since the terms become zero once the argument to  $p$  becomes negative. For example, Figure 2.19 illustrates the calculation of  $p(n)$  from Euler's recursion for  $1 \leq n \leq 12$ . We shall prove Euler's recursion later (see 8.27 and 8.87).

**2.49. Remark: Hardy-Rademacher-Ramanujan Formula for  $p(n)$ .** There exists an

$$\begin{aligned}
p(1) &= p(0) = 1 \\
p(2) &= p(1) + p(0) = 1 + 1 = 2 \\
p(3) &= p(2) + p(1) = 2 + 1 = 3 \\
p(4) &= p(3) + p(2) = 3 + 2 = 5 \\
p(5) &= p(4) + p(3) - p(0) = 5 + 3 - 1 = 7 \\
p(6) &= p(5) + p(4) - p(1) = 7 + 5 - 1 = 11 \\
p(7) &= p(6) + p(5) - p(2) - p(0) = 11 + 7 - 2 - 1 = 15 \\
p(8) &= p(7) + p(6) - p(3) - p(1) = 15 + 11 - 3 - 1 = 22 \\
p(9) &= p(8) + p(7) - p(4) - p(2) = 22 + 15 - 5 - 2 = 30 \\
p(10) &= p(9) + p(8) - p(5) - p(3) = 30 + 22 - 7 - 3 = 42 \\
p(11) &= p(10) + p(9) - p(6) - p(4) = 42 + 30 - 11 - 5 = 56 \\
p(12) &= p(11) + p(10) - p(7) - p(5) + p(0) = 56 + 42 - 15 - 7 + 1 = 77.
\end{aligned}$$

**FIGURE 2.19**

Calculating  $p(n)$  using Euler's recursion.

explicit, non-recursive formula for the number of partitions of  $n$ . The formula is

$$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} A_k(n) \sqrt{k} \left[ \frac{d}{dx} \frac{\sinh\left((\pi/k)\sqrt{\frac{2}{3}(x - \frac{1}{24})}\right)}{\sqrt{x - \frac{1}{24}}} \right] \bigg|_{x=n},$$

where

$$A_k(n) = \sum_{\substack{1 \leq h \leq k: \\ \gcd(h,k)=1}} \omega_{h,k} e^{-2\pi i n h/k},$$

and  $\omega_{h,k}$  is a certain complex  $24k$ th root of unity. By estimating  $p(n)$  by the first term of this series, one can deduce the following asymptotic formula for  $p(n)$ :

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left[\pi\sqrt{2n/3}\right].$$

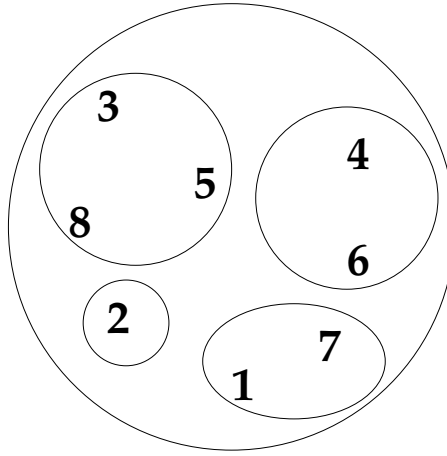
We will not prove these results. For more details, consult Andrews [5, Chapter 5].

## 2.9 Set Partitions

**2.50. Definition: Set Partitions.** Let  $X$  be a set. A *set partition* of  $X$  is a collection  $P$  of nonempty, pairwise disjoint subsets of  $X$  whose union is  $X$ . Each element of  $P$  is called a *block* of the partition. The cardinality of  $P$  (which may be infinite) is called the *number of blocks* of the partition.

For example, if  $X = \{1, 2, 3, 4, 5, 6, 7, 8\}$ , then

$$P = \{\{3, 5, 8\}, \{1, 7\}, \{2\}, \{4, 6\}\}$$

**FIGURE 2.20**

A picture of the set partition  $\{\{3, 5, 8\}, \{1, 7\}, \{2\}, \{4, 6\}\}$ .

is a set partition of  $X$  with four blocks. Note that the ordering of the blocks in this list, and the ordering of the elements within each block, is irrelevant when deciding the equality of two set partitions. For instance,

$$\{\{6, 4\}, \{1, 7\}, \{2\}, \{5, 8, 3\}\}$$

is the same set partition as the partition  $P$  mentioned above. It is convenient to visualize a set partition  $P$  by drawing the elements of  $X$  in a circle, and then drawing smaller circles enclosing the elements of each block of  $P$ . See Figure 2.20.

**2.51. Definition: Stirling Numbers and Bell Numbers.** Let  $S(n, k)$  be the number of set partitions of  $\{1, 2, \dots, n\}$  into exactly  $k$  blocks.  $S(n, k)$  is called a *Stirling number of the second kind*. Let  $B(n)$  be the total number of set partitions of  $\{1, 2, \dots, n\}$ .  $B(n)$  is called a *Bell number*. One can check that  $S(n, k)$  is the number of partitions of *any* given  $n$ -element set into  $k$  blocks; similarly for  $B(n)$ .

Stirling numbers and Bell numbers are not given by closed formulas involving factorials, binomial coefficients, etc. (although there are summation formulas and generating functions for these quantities). However, the Stirling numbers satisfy a recursion that can be used to compute  $S(n, k)$  and  $B(n)$  quite rapidly.

**2.52. Theorem: Recursion for Stirling Numbers of the Second Kind.** For all  $n > 0$  and  $k > 0$ ,

$$S(n, k) = S(n-1, k-1) + kS(n-1, k).$$

The initial conditions are  $S(0, 0) = 1$ ,  $S(n, 0) = 0$  for  $n > 0$ , and  $S(0, k) = 0$  for  $k > 0$ . Furthermore,  $B(0) = 1$  and  $B(n) = \sum_{k=1}^n S(n, k)$  for  $n > 0$ .

*Proof.* Fix  $n, k > 0$ . Let  $A$  be the set of set partitions of  $\{1, 2, \dots, n\}$  into exactly  $k$  blocks. Let  $A' = \{P \in A : \{n\} \in P\}$  and  $A'' = \{P \in A : \{n\} \notin P\}$ .  $A$  is the disjoint union of  $A'$  and  $A''$ .  $A'$  consists of those set partitions such that  $n$  is in a block by itself, while  $A''$  consists of those set partitions such that  $n$  is in a block with some other elements. To build a typical partition  $P \in A'$ , we first choose an arbitrary set partition  $P_0$  of  $\{1, 2, \dots, n-1\}$  into  $k-1$  blocks in any of  $S(n-1, k-1)$  ways. Then we add  $\{n\}$  to  $P_0$  to get  $P$ . To build

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$B(n)$
$n = 0 :$	1	0	0	0	0	0	0	0	0	1
$n = 1 :$	0	1	0	0	0	0	0	0	0	1
$n = 2 :$	0	1	1	0	0	0	0	0	0	2
$n = 3 :$	0	1	3	1	0	0	0	0	0	5
$n = 4 :$	0	1	7	6	1	0	0	0	0	15
$n = 5 :$	0	1	15	25	10	1	0	0	0	52
$n = 6 :$	0	1	31	90	65	15	1	0	0	203
$n = 7 :$	0	1	63	301	350	140	21	1	0	877
$n = 8 :$	0	1	127	966	1701	1050	266	28	1	4140

**FIGURE 2.21**

Calculating  $S(n, k)$  and  $B(n)$  recursively.

a typical partition  $P \in A''$ , we first choose an arbitrary set partition  $P_1$  of  $\{1, 2, \dots, n-1\}$  into  $k$  blocks in any of  $S(n-1, k)$  ways. Then we choose one of these  $k$  blocks and add  $n$  as a new member of that block. By the sum and product rules,

$$S(n, k) = |A| = |A'| + |A''| = S(n-1, k-1) + kS(n-1, k).$$

The initial conditions are immediate from the definitions (note that  $P = \emptyset$  is the unique set partition of  $X = \emptyset$ ). The formula for  $B(n)$  follows from the sum rule.  $\square$

Figure 2.21 computes  $S(n, k)$  and  $B(n)$  for  $n \leq 8$  using the recursion from the last theorem. Note that each entry  $S(n, k)$  in row  $n$  and column  $k$  is computed by taking the number immediately northwest and adding  $k$  times the number immediately above the given entry. The numbers  $B(n)$  are found by adding the numbers in each row.

The Bell numbers also satisfy a nice recursion.

**2.53. Theorem: Recursion for Bell Numbers.** For all  $n > 0$ ,

$$B(n) = \sum_{k=0}^{n-1} \binom{n-1}{k} B(n-1-k).$$

The initial condition is  $B(0) = 1$ .

*Proof.* For  $n > 0$ , we construct a typical set partition  $P$  counted by  $B(n)$  as follows. Let  $k$  be the number of elements in the block of  $P$  containing  $n$ , not including  $n$  itself; thus,  $0 \leq k \leq n-1$ . To build  $P$ , first choose  $k$  elements from  $\{1, 2, \dots, n-1\}$  that belong to the same block as  $n$  in any of  $\binom{n-1}{k}$  ways. Then, choose an arbitrary set partition of the  $n-1-k$  elements that do not belong to the same block as  $n$ ; this choice can be made in any of  $B(n-1-k)$  ways. The recursion now follows from the sum and product rules.  $\square$

For example, assuming that  $B(m)$  is already known for  $m < 8$  (cf. Figure 2.21), we calculate

$$\begin{aligned}
 B(8) &= \binom{7}{0}B(7) + \binom{7}{1}B(6) + \binom{7}{2}B(5) + \cdots + \binom{7}{7}B(0) \\
 &= 1 \cdot 877 + 7 \cdot 203 + 21 \cdot 52 + 35 \cdot 15 + 35 \cdot 5 + 21 \cdot 2 + 7 \cdot 1 + 1 \cdot 1 \\
 &= 4140.
 \end{aligned}$$



We close this section by reviewing the connection between set partitions and equivalence relations.

**2.54. Definition: Types of Relations.** Let  $X$  be any set. A *relation* on  $X$  is any subset of  $X \times X$ . If  $R$  is a relation on  $X$  and  $x, y \in X$ , we often write  $xRy$  as an abbreviation for  $(x, y) \in R$ . We read this symbol as “ $x$  is related to  $y$  under  $R$ .” A relation  $R$  on  $X$  is *reflexive* on  $X$  iff  $xRx$  for all  $x \in X$ .  $R$  is *irreflexive* on  $X$  iff  $xRx$  is false for all  $x \in X$ .  $R$  is *symmetric* iff for all  $x, y \in X$ ,  $xRy$  implies  $yRx$ .  $R$  is *antisymmetric* iff for all  $x, y \in X$ ,  $xRy$  and  $yRx$  imply  $x = y$ .  $R$  is *transitive* iff for all  $x, y, z \in X$ ,  $xRy$  and  $yRz$  imply  $xRz$ .  $R$  is an *equivalence relation* on  $X$  iff  $R$  is symmetric, transitive, and reflexive on  $X$ . If  $R$  is an equivalence relation and  $x_0 \in X$ , the *equivalence class of  $x_0$  relative to  $R$*  is the set  $[x_0]_R = \{y \in X : yRx_0\}$ .

**2.55. Theorem: Set Partitions vs. Equivalence Relations.** Suppose  $X$  is a fixed set. Let  $\mathcal{A}$  be the set of all set partitions of  $X$ , and let  $\mathcal{B}$  be the set of all equivalence relations on  $X$ . There are canonical bijections  $\phi : \mathcal{A} \rightarrow \mathcal{B}$  and  $\phi' : \mathcal{B} \rightarrow \mathcal{A}$ . If  $P \in \mathcal{A}$ , then the number of blocks of  $P$  equals the number of equivalence classes of  $\phi(P)$ . Hence,  $S(n, k)$  is the number of equivalence relations on an  $n$ -element set having  $k$  equivalence classes, and  $B(n)$  is the number of equivalence relations on an  $n$ -element set.

*Proof.* We sketch the proof, leaving certain details as exercises. Given a set partition  $P \in \mathcal{A}$ , define

$$\phi(P) = \{(x, y) \in X : \exists S \in P, x \in S \text{ and } y \in S\}.$$

In other words,  $x\phi(P)y$  iff  $x$  and  $y$  belong to the same block of  $P$ . The reader should check that  $\phi(P)$  is indeed an equivalence relation on  $X$ , i.e., that  $\phi(P) \in \mathcal{B}$ . Thus,  $\phi$  is a well-defined function from  $\mathcal{A}$  into  $\mathcal{B}$ .

Given an equivalence relation  $R \in \mathcal{B}$ , define

$$\phi'(R) = \{[x]_R : x \in X\}.$$

In other words, the blocks of  $\phi'(R)$  are precisely the equivalence classes of  $R$ . The reader should check that  $\phi'(R)$  is indeed a set partition of  $X$ , i.e., that  $\phi'(R) \in \mathcal{A}$ . Thus,  $\phi'$  is a well-defined function from  $\mathcal{B}$  into  $\mathcal{A}$ .

To complete the proof, the reader should check that  $\phi$  and  $\phi'$  are two-sided inverses of one another. In other words, prove that for all  $P \in \mathcal{A}$ ,  $\phi'(\phi(P)) = P$ ; and for all  $R \in \mathcal{B}$ , prove that  $\phi(\phi'(R)) = R$ . It follows that  $\phi$  and  $\phi'$  are bijections.  $\square$

## 2.10 Surjections

Recall that a function  $f : X \rightarrow Y$  is a *surjection* iff for every  $y \in Y$ , there exists  $x \in X$  with  $f(x) = y$ .

**2.56. Definition:**  $\text{Surj}(n, k)$ . Let  $\text{Surj}(n, k)$  denote the number of surjections from an  $n$ -element set onto a  $k$ -element set.

**2.57. Theorem: Recursion for Surjections.** For  $n \geq k > 0$ ,

$$\text{Surj}(n, k) = k \text{Surj}(n-1, k-1) + k \text{Surj}(n-1, k).$$

The initial conditions are  $\text{Surj}(n, k) = 0$  for  $n < k$ ,  $\text{Surj}(0, 0) = 1$ , and  $\text{Surj}(n, 0) = 0$  for  $n > 0$ .

*Proof.* Fix  $n \geq k > 0$ . Let us build a typical surjection  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, k\}$  by considering two cases. Case 1:  $f(i) \neq f(n)$  for all  $i < n$ . In this case, we first choose  $f(n)$  in  $k$  ways, and then we choose a surjection from  $\{1, 2, \dots, n-1\}$  onto  $\{1, 2, \dots, k\} \sim \{f(n)\}$  in  $\text{Surj}(n-1, k-1)$  ways. The total number of possibilities is  $k \text{Surj}(n-1, k-1)$ .

Case 2:  $f(n) = f(i)$  for some  $i < n$ . In this case, note that the restriction of  $f$  to  $\{1, 2, \dots, n-1\}$  is still surjective. Thus we can build  $f$  by first choosing a surjection from  $\{1, 2, \dots, n-1\}$  onto  $\{1, 2, \dots, k\}$  in  $\text{Surj}(n-1, k)$  ways, and then choosing  $f(n) \in \{1, 2, \dots, k\}$  in  $k$  ways. The total number of possibilities is  $k \text{Surj}(n-1, k)$ . The recursion now follows from the sum rule.

The initial conditions are immediate, once we note that the function with graph  $\emptyset$  is the unique surjection from  $\emptyset$  onto  $\emptyset$ .  $\square$

Surjections are closely related to Stirling numbers of the second kind. Indeed, we have the following relation between  $\text{Surj}(n, k)$  and  $S(n, k)$ .

**2.58. Theorem.** For all  $n, k \geq 0$ ,  $\text{Surj}(n, k) = k!S(n, k)$ .

*Proof.* We give two proofs. *First Proof:* We argue by induction on  $n$ . The result holds when  $n = 0$  and  $k$  is arbitrary, since  $\text{Surj}(0, k) = \chi(k = 0) = 0!S(0, k)$ . Assume that  $n > 0$  and that  $\text{Surj}(m, k) = k!S(m, k)$  for all  $k$  and all  $m < n$ . Using the recursions for  $\text{Surj}(n, k)$  and  $S(n, k)$ , we compute

$$\begin{aligned} \text{Surj}(n, k) &= k \text{Surj}(n-1, k-1) + k \text{Surj}(n-1, k) \\ &= k(k-1)!S(n-1, k-1) + k(k!)S(n-1, k) \\ &= k![S(n-1, k-1) + kS(n-1, k)] \\ &= k!S(n, k). \end{aligned}$$

*Second Proof:* We prove the formula by a direct counting argument. To construct a surjection  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, k\}$ , first choose a set partition  $P$  of  $\{1, 2, \dots, n\}$  into  $k$  blocks in any of  $S(n, k)$  ways. Choose one of these blocks (in  $k$  ways), and let  $f$  map everything in this block to 1. Then choose a different block (in  $k-1$  ways), and let  $f$  map everything in this block to 2. Continue similarly; at the last stage, there is 1 block left, and we let  $f$  map everything in this block to  $k$ . By the product rule,

$$\text{Surj}(n, k) = S(n, k) \cdot k \cdot (k-1) \cdot \dots \cdot 1 = k!S(n, k). \quad \square$$

**2.59. Example.** To illustrate the second proof, suppose  $n = 8$  and  $k = 4$ . In the first step, let us choose the partition  $P = \{\{1, 4, 7\}, \{2\}, \{3, 8\}, \{5, 6\}\}$ . In the next four steps, we choose a permutation of the four blocks of  $P$ , say

$$\{2\}, \{5, 6\}, \{3, 8\}, \{1, 4, 7\}.$$

Now we define the associated surjection  $f$  by setting

$$f(2) = 1, f(5) = f(6) = 2, f(3) = f(8) = 3, f(1) = f(4) = f(7) = 4.$$

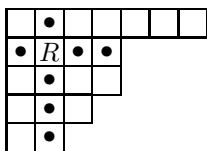
## 2.11 Stirling Numbers and Rook Theory

Recall that the Stirling numbers of the second kind (denoted  $S(n, k)$ ) count the number of set partitions of an  $n$ -element set into  $k$  blocks. This section gives another combinatorial

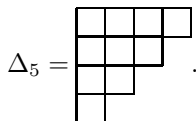
interpretation of these Stirling numbers. We show that  $S(n, k)$  counts certain placements of rooks on a triangular chessboard. A slight variation of this setup leads us to introduce the (signless) Stirling numbers of the first kind. The relationship between the two kinds of Stirling numbers will be illuminated in the following section.

**2.60. Definition: Ferrers Boards and Rooks.** A *Ferrers board* is the diagram of an integer partition, viewed as a collection of unit squares as in §2.8. A *rook* is a chess piece that can occupy any of the squares in a Ferrers board. In chess, a rook can move any number of squares horizontally or vertically from its current position in a single move. A rook located in row  $i$  and column  $j$  of a Ferrers board *attacks* all squares in row  $i$  and all squares in column  $j$ .

For example, in the Ferrers board shown below, the rook  $R$  attacks all squares on the board marked with a dot (and its own square).

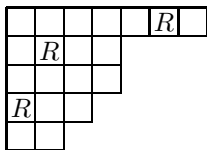


For each  $n > 0$ , let  $\Delta_n$  denote the diagram of the partition  $(n-1, n-2, \dots, 3, 2, 1)$ .  $\Delta_n$  is a triangular Ferrers board with  $n(n-1)/2$  total squares. For example,



**2.61. Definition: Non-attacking Rook Placements.** A *placement* of  $k$  rooks on a given Ferrers board is a subset of  $k$  squares in the Ferrers board. These  $k$  squares represent the locations of  $k$  identical rooks on the board. A placement of rooks in a Ferrers board is called *non-attacking* iff no rook occupies a square attacked by another rook. Equivalently, all rooks in the placement occupy distinct rows and distinct columns of the board.

**2.62. Example.** The following diagram illustrates a non-attacking placement of 3 rooks on the Ferrers board corresponding to the partition  $(7, 4, 4, 3, 2)$ .



**2.63. Theorem: Rook-Theoretic Interpretation of Stirling Numbers of the Second Kind.** For  $n > 0$  and  $0 \leq k \leq n$ , let  $S'(n, k)$  denote the number of non-attacking placements of  $n-k$  rooks on the Ferrers board  $\Delta_n$ . If  $n > 1$  and  $0 < k < n$ , then

$$S'(n, k) = S'(n-1, k-1) + kS'(n-1, k).$$

The initial conditions are  $S'(n, 0) = 0$  and  $S'(n, n) = 1$  for all  $n > 0$ . Therefore,  $S'(n, k) = S(n, k)$ , a Stirling number of the second kind.

*Proof.* Fix  $n > 1$  with  $0 < k < n$ . Let  $A, B, C$  denote the set of placements counted by  $S'(n, k)$ ,  $S'(n-1, k-1)$ , and  $S'(n-1, k)$ , respectively. Let  $A_0$  consist of all rook placements in  $A$  with no rook in the top row, and let  $A_1$  consist of all rook placements in  $A$  with one

rook in the top row.  $A$  is the disjoint union of  $A_0$  and  $A_1$ . Deleting the top row of the Ferrers board  $\Delta_n$  produces the smaller Ferrers board  $\Delta_{n-1}$ . It follows that deleting the (empty) top row of a rook placement in  $A_0$  gives a bijection between  $A_0$  and  $B$  (note that a placement in  $B$  involves  $(n-1) - (k-1) = n-k$  rooks). On the other hand, we can build a typical rook placement in  $A_1$  as follows. First, choose a placement of  $n-k-1$  non-attacking rooks from the set  $C$ , and use this rook placement to fill the bottom  $n-1$  rows of  $\Delta_n$ . These rooks occupy  $n-k-1$  distinct columns. This leaves  $(n-1) - (n-k-1) = k$  columns in the top row in which we are allowed to place the final rook. By the product rule,  $|A_1| = |C|k$ . We conclude that

$$S'(n, k) = |A| = |A_0| + |A_1| = |B| + k|C| = S'(n-1, k-1) + kS'(n-1, k).$$

We cannot place  $n$  non-attacking rooks on the Ferrers board  $\Delta_n$  (which has only  $n-1$  columns), and hence  $S'(n, 0) = 0$ . On the other hand, for any  $n > 0$  there is a unique placement of zero rooks on  $\Delta_n$ . This placement is non-attacking (vacuously), and hence  $S'(n, n) = 1$ . Counting set partitions, we see that  $S(n, 0) = 0$  and  $S(n, n) = 1$  for all  $n > 0$ . Since  $S'(n, k)$  and  $S(n, k)$  satisfy the same recursion and initial conditions, a routine induction argument (cf. 2.34) shows that  $S'(n, k) = S(n, k)$  for all  $n$  and  $k$ .  $\square$

**2.64. Remark.** We have given combinatorial proofs that the numbers  $S'(n, k)$  and  $S(n, k)$  satisfy the same recursion. We can link together these proofs to get a recursively defined bijection between rook placements and set partitions, using the ideas in 2.39. We can also directly define a bijection between rook placements and set partitions. We illustrate such a bijection via an example. Figure 2.22 displays a rook placement counted by  $S'(8, 3)$ . We write the numbers 1 through  $n$  below the last square in each column of the diagram, as shown in the figure. We view these numbers as labeling both the rows and columns of the diagram; note that the column labels increase from left to right, while row labels decrease from top to bottom. The bijection between non-attacking rook placements  $\pi$  and set partitions  $P$  acts as follows. For all  $j < i \leq n$ , there is a rook in row  $i$  and column  $j$  of  $\pi$  iff  $i$  and  $j$  are consecutive elements in the same block of  $P$  (when the elements of the block are written in increasing order). For example, the rook placement  $\pi$  in Figure 2.22 maps to the set partition

$$P = \{\{1, 3, 4, 5, 7\}, \{2, 6\}, \{8\}\}.$$

The set partition  $\{\{2\}, \{1, 5, 8\}, \{4, 6, 7\}, \{3\}\}$  maps to the rook placement shown in Figure 2.23.

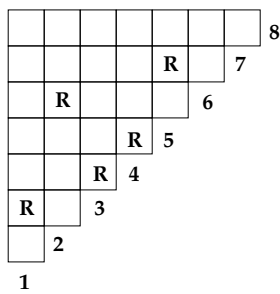
One may check that a non-attacking placement of  $n-k$  rooks on  $\Delta_n$  corresponds to a set partition of  $n$  with exactly  $k$  blocks; furthermore, the rook placement associated to a given set partition is automatically non-attacking.

**2.65. Definition: Wrooks and Stirling Numbers of the First Kind.** A *wrook* (weak rook) is a new chess piece that attacks only the squares in its row. For all  $n > 0$  and  $0 \leq k \leq n$ , let  $s'(n, k)$  denote the number of placements of  $n-k$  non-attacking wrooks on the Ferrers board  $\Delta_n$ . The numbers  $s'(n, k)$  are called *signless Stirling numbers of the first kind*. The numbers  $s(n, k) = (-1)^{n-k}s'(n, k)$  are called (*signed*) *Stirling numbers of the first kind*. Another combinatorial definition of the Stirling numbers of the first kind will be given in §3.6. By convention, we set  $s(0, 0) = 1 = s'(0, 0)$ .

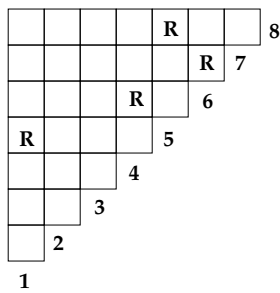
**2.66. Theorem: Recursion for Signless Stirling Numbers of the First Kind.** If  $n > 1$  and  $0 < k < n$ , then

$$s'(n, k) = s'(n-1, k-1) + (n-1)s'(n-1, k).$$

The initial conditions are  $s'(n, 0) = \chi(n=0)$  and  $s'(n, n) = 1$ .



**FIGURE 2.22**  
A rook placement counted by  $S'(n, k)$ , where  $n = 8$  and  $k = 3$ .



**FIGURE 2.23**  
The rook placement associated to  $\{\{2\}, \{1, 5, 8\}, \{4, 6, 7\}, \{3\}\}$ .

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
$n = 0 :$	1	0	0	0	0	0	0	0
$n = 1 :$	0	1	0	0	0	0	0	0
$n = 2 :$	0	-1	1	0	0	0	0	0
$n = 3 :$	0	2	-3	1	0	0	0	0
$n = 4 :$	0	-6	11	-6	1	0	0	0
$n = 5 :$	0	24	-50	35	-10	1	0	0
$n = 6 :$	0	-120	274	-225	85	-15	1	0
$n = 7 :$	0	720	-1764	1624	-735	175	-21	1

**FIGURE 2.24**  
Signed Stirling numbers of the first kind.

*Proof.* Fix  $n > 1$  with  $0 < k < n$ . Let  $A, B, C$  denote the set of placements counted by  $s'(n, k)$ ,  $s'(n-1, k-1)$ , and  $s'(n-1, k)$ , respectively. Write  $A$  as the disjoint union of  $A_0$  and  $A_1$ , where  $A_i$  consists of all elements of  $A$  with  $i$  wrooks in the top row. As above, deletion of the empty top row gives a bijection from  $A_0$  to  $B$ . On the other hand, we can build a typical wrook placement in  $A_1$  as follows. First, choose the position of the wrook in the top row of  $\Delta_n$  in  $n-1$  ways. Second, choose any placement of  $n-k-1$  non-attacking wrooks from the set  $C$ , and use this wrook placement to fill the bottom  $n-1$  rows of  $\Delta_n$ . These wrooks do not attack the wrook in the first row. By the sum and product rules,

$$s'(n, k) = |A| = |A_0| + |A_1| = |B| + (n-1)|C| = s'(n-1, k-1) + (n-1)s'(n-1, k). \quad \square$$

We can use the recursion and initial conditions to compute the (signed or unsigned) Stirling numbers of the first kind. See Figure 2.24, and compare to the computation of Stirling numbers of the second kind in Figure 2.21. There is a surprising relation between the two arrays of numbers in these figures. Specifically, for any fixed  $n > 0$ , consider the lower-triangular matrices  $A = (s(i, j))_{1 \leq i, j \leq n}$  and  $B = (S(i, j))_{1 \leq i, j \leq n}$ . It turns out that  $A$  and  $B$  are inverse matrices! The reader may check this for small  $n$  using Figure 2.21 and Figure 2.24. We will prove this fact for all  $n$  in §2.13.

## 2.12 Linear Algebra Review

In the next few sections, and at other places later in the book, we will need to use some concepts from linear algebra such as vector spaces, bases, and linear independence. This section quickly reviews the definitions we will need; for a thorough treatment of linear algebra, the reader may consult texts such as Hoffman and Kunze [69].

**2.67. Definition: Vector Spaces.** Given a field  $F$ , a *vector space over  $F$*  consists of a set  $V$ , an addition operation  $+: V \times V \rightarrow V$ , and a scalar multiplication operation  $\cdot: F \times V \rightarrow V$ , that satisfy the following axioms.

$\forall x, y \in V, x + y \in V$	(closure under addition)
$\forall x, y, z \in V, x + (y + z) = (x + y) + z$	(associativity of addition)
$\forall x, y \in V, x + y = y + x$	(commutativity of addition)
$\exists 0_V \in V, \forall x \in V, x + 0_V = x = 0_V + x$	(existence of additive identity)
$\forall x \in V, \exists -x \in V, x + (-x) = 0_V = (-x) + x$	(existence of additive inverses)
$\forall c \in F, \forall v \in V, c \cdot v \in V$	(closure under scalar multiplication)
$\forall c \in F, \forall v, w \in V, c \cdot (v + w) = (c \cdot v) + (c \cdot w)$	(left distributive law)
$\forall c, d \in F, \forall v \in V, (c + d) \cdot v = (c \cdot v) + (d \cdot v)$	(right distributive law)
$\forall c, d \in F, \forall v \in V, (cd) \cdot v = c \cdot (d \cdot v)$	(associativity of scalar multiplication)
$\forall v \in V, 1 \cdot v = v$	(identity property)

When discussing vector spaces, elements of  $V$  are often called *vectors*, while elements of  $F$  are called *scalars*.

For example, the set  $F^n = \{(x_1, \dots, x_n) : x_i \in F\}$  is a vector space over  $F$  with operations

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n);$$

$$c(x_1, \dots, x_n) = (cx_1, \dots, cx_n) \quad (c, x_i, y_i \in F).$$

Similarly, the set of polynomials  $a_0 + a_1x + \cdots + a_kx^k$ , where all coefficients  $a_i$  come from  $F$ , is a vector space over  $F$  under the operations

$$\sum_{i \geq 0} a_i x^i + \sum_{i \geq 0} b_i x^i = \sum_{i \geq 0} (a_i + b_i) x^i; \quad c \sum_{i \geq 0} a_i x^i = \sum_{i \geq 0} (ca_i) x^i.$$

We consider two polynomials  $\sum_{i \geq 0} a_i x^i$  and  $\sum_{i \geq 0} b_i x^i$  to be *equal* iff  $a_i = b_i$  for all  $i$ ; see §7.3 for a more formal discussion of how to define polynomials.

**2.68. Definition: Spanning Sets and Linear Combinations.** A subset  $S$  of a vector space  $V$  over  $F$  *spans*  $V$  iff for every  $v \in V$ , there exists a *finite* list of vectors  $v_1, \dots, v_k \in S$  and scalars  $c_1, \dots, c_k \in F$  with  $v = c_1 v_1 + \cdots + c_k v_k$ . Any expression of the form  $c_1 v_1 + \cdots + c_k v_k$  is called a *linear combination* of  $v_1, \dots, v_k$ . A linear combination must be a *finite* sum of vectors.

**2.69. Definition: Linear Independence.** A list  $(v_1, \dots, v_k)$  of vectors in a vector space  $V$  over  $F$  is called *linearly dependent* iff there exist scalars  $c_1, \dots, c_k \in F$  such that  $c_1 v_1 + \cdots + c_k v_k = 0_V$  and at least one  $c_i$  is not zero. Otherwise, the list  $(v_1, \dots, v_k)$  is called *linearly independent*. A set  $S \subseteq V$  (possibly infinite) is *linearly dependent* iff there is a finite list of *distinct* elements of  $S$  that is linearly dependent; otherwise,  $S$  is *linearly independent*.

**2.70. Definition: Basis of a Vector Space.** A *basis* of a vector space  $V$  is a set  $S \subseteq V$  that is linearly independent and spans  $V$ .

For example, for any field  $F$ , define  $e_i \in F^n$  to be the vector with  $1_F$  in position  $i$  and  $0_F$  in all other positions. Then  $\{e_1, \dots, e_n\}$  is a basis for  $F^n$ . Similarly, one may check that the infinite set  $S = \{1, x, x^2, x^3, \dots, x^n, \dots\}$  is a basis for the vector space  $V$  of polynomials in  $x$  with coefficients in  $F$ .  $S$  spans  $V$  since every polynomial must be a *finite* linear combination of powers of  $x$ . The linear independence of  $S$  follows from the definition of polynomial equality: the only linear combination  $c_0 1 + c_1 x + c_2 x^2 + \cdots$  that can equal the zero polynomial is the one where  $c_0 = c_1 = c_2 = \cdots = 0_F$ . We now state without proof some of the fundamental facts about spanning sets, linear independence, and bases.

**2.71. Theorem: Linear Algebra Facts.** Every vector space  $V$  over a field  $F$  has a basis (possibly infinite). Any two bases of  $V$  have the same cardinality, which is called the *dimension* of  $V$  and denoted  $\dim(V)$ . Given a basis of  $V$ , every  $v \in V$  can be expressed in exactly one way as a linear combination of the basis elements. Any linearly independent set in  $V$  can be enlarged to a basis of  $V$ . Any spanning set for  $V$  contains a basis of  $V$ . A set  $S \subseteq V$  with  $|S| > \dim(V)$  must be linearly dependent. A set  $T \subseteq V$  with  $|T| < \dim(V)$  cannot span  $V$ .

For example,  $\dim(F^n) = n$  for all  $n \geq 1$ , whereas the vector space of polynomials with coefficients in  $F$  has (countably) infinite dimension.

## 2.13 Stirling Numbers and Polynomials

In this section, we use our recursions for Stirling numbers (of both kinds) to give algebraic proofs of certain polynomial identities. We will see that these identities connect certain frequently used bases of the vector space of one-variable polynomials. This linear-algebraic interpretation of Stirling numbers will be used to show the inverse relation between the two triangular matrices of Stirling numbers (cf. Figures 2.21 and 2.24). The following section will give combinatorial proofs of the same identities using rook theory.

**2.72. Theorem: Polynomial Identity for Stirling Numbers of the Second Kind.**

For all  $n \geq 0$  and all real  $x$ ,

$$x^n = \sum_{k=0}^n S(n, k)x(x-1)(x-2)\cdots(x-k+1). \quad (2.6)$$

*Proof.* We give an algebraic proof here; the next section gives a combinatorial proof using rook theory. Recall that  $S(0, 0) = 1$  and  $S(n, k) = S(n-1, k-1) + kS(n-1, k)$  for  $n \geq 1$  and  $1 \leq k \leq n$ . We prove the stated identity by induction on  $n$ . If  $n = 0$ , the right side is  $S(0, 0) = 1 = x^0$ , so the identity holds. For the induction step, fix  $n \geq 1$  and assume that  $x^{n-1} = \sum_{k=0}^{n-1} S(n-1, k)x(x-1)\cdots(x-k+1)$ . Multiplying both sides by  $x = (x-k) + k$ , we can write

$$\begin{aligned} x^n &= \sum_{k=0}^{n-1} S(n-1, k)x(x-1)\cdots(x-k+1)(x-k) \\ &\quad + \sum_{k=0}^{n-1} S(n-1, k)x(x-1)\cdots(x-k+1)k \\ &= \sum_{j=0}^{n-1} S(n-1, j)x(x-1)\cdots(x-j) + \sum_{k=0}^{n-1} kS(n-1, k)x(x-1)\cdots(x-k+1). \end{aligned}$$

In the first summation, replace  $j$  by  $k-1$ . The calculation continues:

$$\begin{aligned} x^n &= \sum_{k=1}^n S(n-1, k-1)x(x-1)\cdots(x-k+1) \\ &\quad + \sum_{k=0}^{n-1} kS(n-1, k)x(x-1)\cdots(x-k+1) \\ &= \sum_{k=0}^n S(n-1, k-1)x(x-1)\cdots(x-k+1) \\ &\quad + \sum_{k=0}^n kS(n-1, k)x(x-1)\cdots(x-k+1) \quad (\text{since } S(n-1, -1) = S(n-1, n) = 0) \\ &= \sum_{k=0}^n [S(n-1, k-1) + kS(n-1, k)]x(x-1)\cdots(x-k+1) \\ &= \sum_{k=0}^n S(n, k)x(x-1)\cdots(x-k+1) \quad (\text{using the recursion for } S(n, k)). \quad \square \end{aligned}$$

**2.73. Theorem: Polynomial Identity for Signless Stirling Numbers of the First Kind.**

For all  $n \geq 0$  and all real  $x$ ,

$$x(x+1)(x+2)\cdots(x+n-1) = \sum_{k=0}^n s'(n, k)x^k. \quad (2.7)$$

*Proof.* Recall that  $s'(0, 0) = 1$  and  $s'(n, k) = s'(n-1, k-1) + (n-1)s'(n-1, k)$  for  $n \geq 1$  and  $1 \leq k \leq n$ . We use induction on  $n$  again. If  $n = 0$ , both sides of (2.7) evaluate to 1. For the induction step, fix  $n \geq 1$  and assume that

$$x(x+1)(x+2)\cdots(x+n-2) = \sum_{k=0}^{n-1} s'(n-1, k)x^k.$$



Multiply both sides of this assumption by  $x + n - 1$  and compute:

$$\begin{aligned}
 x(x+1)\cdots(x+n-1) &= \sum_{k=0}^{n-1} s'(n-1, k)x^k(x+n-1) \\
 &= \sum_{k=0}^{n-1} s'(n-1, k)x^{k+1} + \sum_{k=0}^{n-1} (n-1)s'(n-1, k)x^k \\
 &= \sum_{k=1}^n s'(n-1, k-1)x^k + \sum_{k=0}^{n-1} (n-1)s'(n-1, k)x^k \\
 &= \sum_{k=0}^n [s'(n-1, k-1) + (n-1)s'(n-1, k)]x^k \\
 &= \sum_{k=0}^n s'(n, k)x^k \quad (\text{using the recursion for } s'(n, k)). \quad \square
 \end{aligned}$$

**2.74. Theorem: Polynomial Identity for Signed Stirling Numbers of the First Kind.** For all  $n \geq 0$  and all real  $x$ ,

$$x(x-1)(x-2)\cdots(x-n+1) = \sum_{k=0}^n s(n, k)x^k. \quad (2.8)$$

*Proof.* Replace  $x$  by  $-x$  in (2.7) to obtain

$$(-x)(-x+1)(-x+2)\cdots(-x+n-1) = \sum_{k=0}^n s'(n, k)(-x)^k.$$

Factoring out  $-1$ 's, we get

$$(-1)^n x(x-1)(x-2)\cdots(x-n+1) = \sum_{k=0}^n (-1)^k s'(n, k)x^k.$$

Moving the  $(-1)^n$  to the right side and recalling that  $s(n, k) = (-1)^{n+k} s'(n, k)$ , the result follows.  $\square$

**2.75. Theorem: Summation Formulas for Stirling Numbers of the First Kind.** For all  $n \geq 1$  and  $1 \leq k \leq n$ , we have

$$s'(n, k) = \sum_{1 \leq i_1 < i_2 < \cdots < i_{n-k} \leq n-1} i_1 i_2 \cdots i_{n-k}.$$

*Proof.* We give an algebraic proof and a combinatorial proof of this result.

*Algebraic Proof.* We apply the generalized distributive law to the left side of the identity

$$(x+0)(x+1)(x+2)\cdots(x+n-1) = \sum_{k=0}^n s'(n, k)x^k.$$

According to the distributive law, the left side expands to a sum of terms obtained by choosing either  $x$  or  $i$  from each factor (for  $0 \leq i < n$ ) and multiplying the chosen terms together. To obtain a contribution to the coefficient of  $x^k$ , we must choose  $x$  exactly  $k$  times and choose a number  $i$  exactly  $n - k$  times. Adding up all these contributions gives the

coefficient of  $x^k$ , namely  $s'(n, k)$ . The term  $i_1 i_2 \cdots i_{n-k}$  is the contribution from the choice sequence where we choose  $i_1$  from the factor  $(x + i_1)$ ,  $i_2$  from the factor  $(x + i_2)$ , etc., and choose  $x$  from all factors  $(x + i)$  with  $i$  different from all  $i_j$ 's.

*Combinatorial Proof.* Recall that  $s'(n, k)$  counts the number of placements of  $n - k$  non-attacking wrooks on the triangular Ferrers board  $\Delta_n$ . Since wrooks only attack cells in their rows, a placement is non-attacking iff all wrooks occupy distinct rows of  $\Delta_n$ . Let us classify wrook placements based on which rows contain wrooks. Suppose the  $n - k$  wrooks appear in the rows of lengths  $i_1, i_2, \dots, i_{n-k}$ , where  $1 \leq i_1 < i_2 < \cdots < i_{n-k} \leq n - 1$ . The product rule shows that the number of placements of wrooks in these rows is  $i_1 i_2 \cdots i_{n-k}$ . The formula in the theorem now follows from the sum rule.  $\square$

**2.76. Definition: Special Bases for Polynomials.** Let  $V$  be the vector space of all polynomials in one variable  $x$  with real coefficients. For any integer  $n \geq 0$ , introduce the *falling factorial* polynomials

$$(x)_{\downarrow 0} = 1, \quad (x)_{\downarrow n} = x(x-1)(x-2) \cdots (x-n+1).$$

Similarly, the *rising factorial* polynomials are defined by

$$(x)_{\uparrow 0} = 1, \quad (x)_{\uparrow n} = x(x+1)(x+2) \cdots (x+n-1).$$

The *monomial basis* of  $V$  is the indexed set  $M = \{x^n : n \geq 0\}$ . The *falling factorial basis* of  $V$  is  $F = \{(x)_{\downarrow n} : n \geq 0\}$ . The *rising factorial basis* of  $V$  is  $R = \{(x)_{\uparrow n} : n \geq 0\}$ .

It is a routine exercise to prove that any indexed collection of polynomials  $\{p_n(x) : n \geq 0\}$  such that  $\deg(p_n) = n$  for all  $n$  is a basis of  $V$ . Since  $x^n$ ,  $(x)_{\downarrow n}$ , and  $(x)_{\uparrow n}$  all have degree  $n$ , it follows that  $M$ ,  $F$ , and  $R$  really are bases of  $V$ . Define  $M_N = \{x^n : 0 \leq n \leq N\}$ , and define  $F_N$  and  $R_N$  similarly. The three indexed collections  $M_N$ ,  $F_N$ , and  $R_N$  are all bases of the vector space  $V_N$  of polynomials in  $x$  of degree at most  $N$ .

We can now recast the preceding theorems in the language of linear algebra. Recall that if  $B = (v_1, \dots, v_n)$  and  $C = (w_1, \dots, w_n)$  are two ordered bases of a finite-dimensional vector space  $W$ , the *transition matrix from  $B$  to  $C$*  is the unique  $n \times n$  matrix  $A = (a_{ij})$  such that

$$v_j = \sum_{i=1}^n a_{ij} w_i \quad (1 \leq j \leq n). \quad (2.9)$$

This matrix is so named because if  $v \in W$  has coordinates  $[v]_B = (s_1, \dots, s_n)^T$  relative to the basis  $B$  (i.e.,  $v = \sum_j s_j v_j$ ), then the coordinates of  $v$  relative to the basis  $C$  are given by  $[v]_C = A[v]_B$ . Thus, multiplication by  $A$  transforms coordinates relative to  $B$  into coordinates relative to  $C$ . From linear algebra, we know that  $A$  is invertible, and  $A^{-1}$  is none other than the transition matrix from  $C$  to  $B$ .

**2.77. Theorem: Transition Matrices between Polynomial Bases.** Fix  $N \geq 0$ , and write  $M_N = (x^n : n \leq N)$ ,  $F_N = ((x)_{\downarrow n} : n \leq N)$ , and  $R_N = ((x)_{\uparrow n} : n \leq N)$ , as above.

(a) The matrix  $\mathbf{S} = (S(n, k))_{0 \leq n, k \leq N}$  of Stirling numbers of the second kind is the transpose of the transition matrix from the basis  $M_N$  to the basis  $F_N$  of the vector space  $V_N$  of polynomials of degree at most  $N$ .

(b) The matrix  $\mathbf{s}' = (s'(n, k))_{0 \leq n, k \leq N}$  of signless Stirling numbers of the first kind is the transpose of the transition matrix from the basis  $R_N$  to the basis  $M_N$  of  $V_N$ .

(c) The matrix  $\mathbf{s} = (s(n, k))_{0 \leq n, k \leq N}$  of signed Stirling numbers of the first kind is the transpose of the transition matrix from the basis  $F_N$  to the basis  $M_N$  of  $V_N$ .

(d) The  $(N+1) \times (N+1)$  matrices  $\mathbf{S}$  and  $\mathbf{s}$  are inverses of one another.



and the theorem follows.  $\square$

Comparing the combinatorial proof in 2.78 to the algebraic proof in 2.73, a subtle difference emerges: the combinatorial proof is valid only for *nonnegative integers*  $x$ , while the algebraic proof is valid for all *real*  $x$  (or even for formal polynomials in any polynomial ring  $F[x]$ , as defined in §7.3). The following result shows that our combinatorial proof is equally as good as the algebraic proof.

**2.79. Theorem: Verifying Polynomial Identities.** Suppose  $F$  is a field and  $p, q$  are two polynomials in  $F[x]$ . Say  $p$  and  $q$  have degree at most  $N$ . If  $p(c) = q(c)$  for  $N + 1$  elements  $c \in F$ , then  $p = q$  in the polynomial ring  $F[x]$ , and hence  $p(c) = q(c)$  for all  $c \in F$ . In particular, if two real polynomials agree at each nonnegative integer, then the two polynomials are identical.

*Proof.* Consider the polynomial  $p - q \in F[x]$ . If this polynomial is nonzero, its degree is at most  $N$ . A well-known fact from algebra asserts that a nonzero polynomial of degree at most  $N$  has at most  $N$  roots in  $F$  (cf. 2.157 and 12.147). Our hypothesis says that  $p - q$  has  $N + 1$  roots in  $F$ . Therefore,  $p - q = 0$  in  $F[x]$ . Evaluating  $p - q$  at any field element  $c$ , it follows that  $p(c) = q(c)$ .  $\square$

**2.80. Example.** To apply 2.79 to 2.78, fix  $n$ . The left side of (2.10), namely  $p = x(x + 1) \cdots (x + n - 1)$ , is a polynomial in  $x$  of degree  $n$ . The right side  $q = \sum_{k=0}^n s'(n, k)x^k$  is also a real polynomial in  $x$  of degree  $n$ . The wrook-theoretic proof in 2.78 showed that  $p(m) = q(m)$  for all integers  $m \geq 0$ . Hence,  $p = q \in \mathbb{R}[x]$  and thus  $p(r) = q(r)$  for all real  $r$ .

This type of argument involving 2.79 occurs so commonly that we will seldom spell out the details in the future. However, one must remember to check that both sides of the identity in question are *polynomials* in the variable  $x$ .

**2.81. Theorem: Stirling Numbers of the Second Kind Revisited.** For all integers  $n > 0$  and all real  $x$ ,

$$x^n = \sum_{k=0}^n S(n, k)x(x-1)(x-2) \cdots (x-k+1). \quad (2.11)$$

*Proof.* Both sides of the identity are polynomials in  $x$ , so it suffices to verify the identity when  $x$  is a nonnegative integer. Fix  $x \geq 0$  and  $n > 0$ . Let  $A$  be the set of placements of  $n$  non-attacking rooks on the extended Ferrers board  $\Delta_n(x) = \text{dg}(x + n - 1, x + n - 2, \dots, x + 1, x)$ . We can build a placement  $\pi \in A$  by placing one rook in each row, working from bottom to top. The rook in the bottom row can go in any of  $x$  squares. The rook in the next row can go in any of  $(x + 1) - 1 = x$  squares, since one column is attacked by the rook in the bottom row. In general, the rook located  $i \geq 0$  rows above the bottom row can go in  $(x + i) - i = x$  squares, since  $i$  distinct columns are already attacked by lower rooks when the time comes to place the rook in this row. The product rule therefore gives  $|A| = x^n$ .

On the other hand,  $A$  is the disjoint union of the sets  $A_k$  consisting of all  $\pi \in A$  with exactly  $k$  rooks in the new squares and  $n - k$  rooks in the old squares. (Recall that the new squares are the leftmost  $x$  columns of  $\Delta_n(x)$ .) To build  $\pi \in A_k$ , first place  $n - k$  non-attacking rooks in  $\Delta_n$  in any of  $S(n, k)$  ways. There are now  $k$  unused rows of new squares, each of length  $x$ . Visit these rows from top to bottom (say), placing one rook in each row. There are  $x$  choices for the first rook, then  $x - 1$  choices for the second rook (since the first rook's column must be avoided), then  $x - 2$  choices for the third rook, etc. The product rule gives  $|A_k| = S(n, k)x(x-1)(x-2) \cdots (x-k+1)$ . Hence, by the sum rule,

$$|A| = \sum_{k=1}^n S(n, k)x(x-1)(x-2) \cdots (x-k+1).$$

The result follows by comparing our two formulas for  $|A|$ .  $\square$

**2.82. Remark.** The proof technique of using extended boards such as  $\Delta_n(x)$  can be reused to prove other results in rook theory. See §12.3.

We can also prove polynomial identities in several variables by verifying that the identity holds for sufficiently many values of the variables. As an example, we now present a combinatorial proof of the binomial theorem.

**2.83. Combinatorial Binomial Theorem.** For all integers  $n \geq 0$ ,

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \text{ in } \mathbb{R}[x, y].$$

*Proof.* Let  $p = (x + y)^n \in \mathbb{R}[x, y]$  and  $q = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \in \mathbb{R}[x, y]$ . We first show that  $p(i, j) = q(i, j)$  for all nonnegative integers  $i, j \geq 0$ . Fix such integers  $i$  and  $j$ . Consider an alphabet  $A$  consisting of  $i$  consonants and  $j$  vowels. Let  $B$  be the set of all  $n$ -letter words using letters from  $A$ . The product rule gives  $|B| = (i + j)^n = p(i, j)$ . On the other hand,  $B$  is the disjoint union of sets  $B_k$  (for  $0 \leq k \leq n$ ) where  $B_k$  consists of the words in  $B$  with exactly  $k$  consonants. To build a word  $w \in B_k$ , first choose the positions for the  $k$  consonants out of the  $n$  available positions in  $\binom{n}{k}$  ways. Choose the  $k$  consonants in these positions from left to right ( $i$  ways each), and then choose the  $n - k$  vowels in the remaining positions from left to right ( $j$  ways each). The product rule gives  $|B_k| = \binom{n}{k} i^k j^{n-k}$ . Hence, the sum rule gives  $|B| = q(i, j)$ .

We complete the proof by invoking 2.79 twice. First, for each nonnegative integer  $i \geq 0$ , the polynomials  $p(i, y)$  and  $q(i, y)$  in  $\mathbb{R}[y]$  agree for infinitely many values of the formal variable  $y$ . So,  $p(i, y) = q(i, y)$  in  $\mathbb{R}[y]$  and also in  $\mathbb{R}(y)$  (the field of fractions of the polynomial ring  $\mathbb{R}[y]$ , cf. 7.44). Now regard  $p$  and  $q$  as elements of the polynomial ring  $\mathbb{R}(y)[x]$ . We have just shown that these polynomials (viewed as polynomials in the single variable  $x$ ) agree in  $\mathbb{R}(y)$  for infinitely many values of  $x$ . Hence,  $p = q$  in  $\mathbb{R}(y)[x]$ , and so  $p = q$  in  $\mathbb{R}[x, y]$ . This argument generalizes to polynomial identities in any number of variables (see 2.158), so we will omit the details in the future.  $\square$

## Summary

- *Generalized Distributive Law.* To multiply out a product of factors, where each factor is a sum of terms, choose one term from each factor, multiply these choices together, and add up the resulting products. Formally, this can be written:

$$\prod_{k=1}^n \left( \sum_{i_k \in I_k} x_{k, i_k} \right) = \sum_{(i_1, \dots, i_n) \in I_1 \times \dots \times I_n} \left( \prod_{k=1}^n x_{k, i_k} \right) \quad (\text{all } x_{k, j} \text{ lie in a ring } R).$$

If  $A_1, \dots, A_n$  are finite subsets of  $R$ , we can also write

$$\left( \sum_{w_1 \in A_1} w_1 \right) \cdot \left( \sum_{w_2 \in A_2} w_2 \right) \cdots \left( \sum_{w_n \in A_n} w_n \right) = \sum_{(w_1, w_2, \dots, w_n) \in A_1 \times A_2 \times \dots \times A_n} w_1 w_2 \cdots w_n.$$

- *Multinomial and Binomial Theorems.* In any ring  $R$  (possibly non-commutative),

$$(z_1 + z_2 + \cdots + z_s)^n = \sum_{\text{words } w \in \{1, \dots, s\}^n} z_{w_1} z_{w_2} \cdots z_{w_n} \quad (s, n \in \mathbb{N}^+, z_i \in R).$$

If  $z_i z_j = z_j z_i$  for all  $i, j$ , this becomes

$$(z_1 + z_2 + \cdots + z_s)^n = \sum_{n_1 + n_2 + \cdots + n_s = n} \binom{n}{n_1, n_2, \dots, n_s} z_1^{n_1} z_2^{n_2} \cdots z_s^{n_s}.$$

If  $xy = yx$ , we have the binomial theorem

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$

- *Combinatorial Proofs.* To prove a formula of the form  $a = b$  combinatorially, one can define a set  $S$  of objects and give two counting arguments showing  $|S| = a$  and  $|S| = b$ , respectively. To prove a polynomial identity  $p(x) = q(x)$ , it suffices to verify the identity for infinitely many values of the variable  $x$  (say for all nonnegative integers  $x$ ). Similar comments apply to multivariable polynomial identities.
- *Identities for Binomial Coefficients.*

$$\binom{n}{k} = \binom{n}{n-k}; \quad \sum_{k=0}^n \binom{n}{k} = 2^n; \quad \sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}; \quad \binom{a+b}{a, b} = \sum_{k=0}^a \binom{k+b-1}{k, b-1};$$

$$\binom{a+b+c+1}{a, b+c+1} = \sum_{k=0}^a \binom{k+b}{k, b} \binom{a-k+c}{a-k, c} \quad (\text{Chu-Vandermonde formula})$$

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \quad (\text{Pascal recursion}).$$

- *Combinatorial Definitions.* An *integer partition of  $n$  into  $k$  parts* is a weakly decreasing list  $\mu = (\mu_1, \mu_2, \dots, \mu_k)$  of positive integers that sum to  $n$ . A *set partition of a set  $S$  into  $k$  blocks* is a set  $P$  of  $k$  nonempty, pairwise disjoint sets with union  $S$ . An *equivalence relation on  $S$*  is a reflexive, symmetric, transitive relation on  $S$ . The map that sends an equivalence relation to the set of its equivalence classes defines a bijection from the set of equivalence relations on  $S$  to the set of set partitions of  $S$ .
- *Notation for Combinatorial Objects.* Table 2.1 indicates notation used for counting some collections of combinatorial objects.
- *Recursions.* A collection of combinatorial objects can often be described *recursively*, by using smaller objects of the same kind to build larger objects. Induction arguments can be used to prove facts about recursively defined objects. Table 2.2 lists some recursions satisfied by the quantities in Table 2.1. In each case, the recursion together with appropriate *initial conditions* uniquely determine the quantities under consideration. If two collections of objects satisfy the same recursion and initial conditions, one can link together two combinatorial proofs of the recursion to obtain recursively defined bijections between the two collections.

**TABLE 2.1**

Notation for counting combinatorial objects.

Here  $f_n$  is a Fibonacci number;  $B(n)$  is a Bell number;  $S(n, k)$  is a Stirling number of the second kind;  $s'(n, k)$  is a signless Stirling number of the first kind;  $T(a, b)$  is a ballot number;  $T_m(a, b)$  is an  $m$ -ballot number; and  $C_n$  is a Catalan number.

Notation	What it counts
$C(n, k) = \binom{n}{k}$	$k$ -element subsets of an $n$ -element set
$C(n; n_1, \dots, n_s) = \binom{n}{n_1, \dots, n_s}$	anagrams of the word $a_1^{n_1} \dots a_s^{n_s}$
$L(a, b) = \binom{a+b}{a, b}$	lattice paths from $(0, 0)$ to $(a, b)$
$f_n$	words in $\{0, 1\}^n$ not containing 00
$M(n, k) = \binom{k+n-1}{k, n-1}$	$k$ -element multisets of an $n$ -element set
$B(n)$	set partitions of an $n$ -element set;
$S(n, k)$	equivalence relations on an $n$ -element set
	set partitions of $\{1, \dots, n\}$ into $k$ blocks;
	equiv. relations on $\{1, \dots, n\}$ with $k$ equiv. classes;
$\text{Surj}(n, k) = k!S(n, k)$	placements of $n - k$ non-attacking rooks on $\Delta_n$
$s'(n, k)$	surjections from $\{1, \dots, n\}$ to $\{1, \dots, k\}$
	placements of $n - k$ non-attacking wrooks on $\Delta_n$ ;
	permutations of $n$ objects with $k$ cycles (§3.6)
$s(n, k) = (-1)^{n+k} s'(n, k)$	Stirling numbers of first kind (with signs)
$p(n)$	integer partitions of $n$
$p(n, k)$	integer partitions of $n$ into $k$ parts
$T(a, b) = \frac{b-a+1}{b+a+1} \binom{a+b+1}{a, b+1}$	lattice paths from $(0, 0)$ to $(a, b)$ that
	do not go below $y = x$ ( $b \geq a$ )
$T_m(a, b) = \frac{b-ma+1}{b+a+1} \binom{a+b+1}{a, b+1}$	lattice paths from $(0, 0)$ to $(a, b)$ that
	do not go below $y = mx$ ( $m \in \mathbb{N}^+, b \geq ma$ )
$C_n = T(n, n) = \frac{1}{n+1} \binom{2n}{n, n}$	Dyck paths ending at $(n, n)$ ;
	binary trees with $n$ vertices;
	231-avoiding permutations; etc.

**TABLE 2.2**

Some combinatorial recursions.

$$\begin{aligned}
C(n, k) &= C(n-1, k-1) + C(n-1, k) \\
C(n; n_1, \dots, n_s) &= \sum_{k=1}^s C(n-1; n_1, \dots, n_k-1, \dots, n_s) \\
f_n &= f_{n-1} + f_{n-2} \\
M(n, k) &= M(n-1, k) + M(n, k-1) \\
L(a, b) &= L(a-1, b) + L(a, b-1) \\
T(a, b) &= T(a-1, b) + T(a, b-1)\chi(b-1 \geq a) \\
C_n &= \sum_{k=1}^n C_{k-1}C_{n-k} \\
p(n, k) &= p(n-1, k-1) + p(n-k, k) \\
p(n) &= \sum_{m \geq 1} (-1)^{m-1} \left[ p\left(n - \frac{m(3m-1)}{2}\right) + p\left(n - \frac{m(3m+1)}{2}\right) \right] \\
S(n, k) &= S(n-1, k-1) + kS(n-1, k) \\
B(n) &= \sum_{k=0}^{n-1} \binom{n-1}{k} B(n-1-k) \\
\text{Surj}(n, k) &= k \text{Surj}(n-1, k-1) + k \text{Surj}(n-1, k) \\
s'(n, k) &= s'(n-1, k-1) + (n-1)s'(n-1, k)
\end{aligned}$$

- *Polynomial Identities for Stirling numbers.* Define *rising* and *falling* factorials by  $(x)\uparrow_n = x(x+1)(x+2)\cdots(x+n-1)$  and  $(x)\downarrow_n = x(x-1)(x-2)\cdots(x-n+1)$ . The sets  $\{x^n : n \geq 0\}$ ,  $\{(x)\uparrow_n : n \geq 0\}$ , and  $\{(x)\downarrow_n : n \geq 0\}$  are all bases of the vector space of real polynomials in one variable. The Stirling numbers are the entries in the transition matrices between these bases. More specifically,

$$x^n = \sum_k S(n, k)(x)\downarrow_k; \quad (x)\uparrow_n = \sum_k s'(n, k)x^k; \quad (x)\downarrow_n = \sum_k s(n, k)x^k.$$

So the matrices  $\mathbf{S} = (S(n, k))_{n, k}$  and  $\mathbf{s} = (s(n, k))_{n, k}$  are inverses of each other, i.e.,

$$\sum_k S(i, k)s(k, j) = \chi(i = j) = \sum_k s(i, k)S(k, j) \quad (i, j \geq 0).$$

## Exercises

**2.84.** Simplify the product  $(B + C + H)(A + E + U)(R + T)$ , where each letter denotes an arbitrary  $n \times n$  real matrix.

**2.85.** Expand  $(A + B + C)^3$ , where  $A, B, C$  are  $n \times n$  matrices.



- 2.86.** Find the coefficient of  $w^2x^3yz^3$  in  $(w+x+y+z)^9$ , assuming  $w, x, y$ , and  $z$  commute.
- 2.87.** Expand  $(3x-2)^5$  into a sum of monomials.
- 2.88.** Find the constant term in  $(2x-x^{-1})^6$ .
- 2.89.** Find the coefficient of  $x^3$  in  $(x^2+x+1)^4$ .
- 2.90.** Prove algebraically that  $\sum_{k=0}^n (-1)^k \binom{n}{k} = \chi(n=0)$  and  $\sum_{k=0}^n 2^k \binom{n}{k} = 3^n$  for all  $n \geq 0$ . Can you find combinatorial proofs?
- 2.91.** Given  $n \in \mathbb{N}^+$ , evaluate  $\sum_{0 \leq j < k \leq n} \binom{n}{j} \binom{n}{k}$ .
- 2.92.** Given  $m, n \in \mathbb{N}^+$ , evaluate  $\sum_{k_1+k_2+\dots+k_m=n} (k_1!k_2!\dots k_m!)^{-1}$ .
- 2.93.** Use Pascal's recursion to compute  $\binom{9}{k}$  for  $0 \leq k \leq 9$  and  $\binom{10}{k}$  for  $0 \leq k \leq 10$ .
- 2.94.** Give a proof of the recursion 2.27 for multinomial coefficients based on multidimensional lattice paths.
- 2.95.** Compute the ballot numbers  $T(a, 7)$  for  $0 \leq a \leq 7$  by drawing a picture.
- 2.96.** For fixed  $k \in \mathbb{N}^+$ , let  $a_n$  be the number of  $n$ -letter words using the alphabet  $\{0, 1, \dots, k\}$  that do not contain 00. Find a recursion and initial conditions for  $a_n$ .
- 2.97.** How many words in  $\{0, 1, 2\}^8$  do not contain 000?
- 2.98.** How many lattice paths from  $(1, 1)$  to  $(6, 6)$  always stay weakly between the lines  $y = 2x/5$  and  $y = 5x/2$ ?
- 2.99.** How many lattice paths go from  $(1, 1)$  to  $(8, 8)$  without ever passing through a point  $(p, q)$  such that  $p$  and  $q$  are both prime?
- 2.100.** Show that  $C_n$  counts integer partitions  $\mu$  such that  $\text{dg}(\mu) \subseteq \Delta_n$ .
- 2.101.** Draw pictures of all integer partitions of  $n = 6$  and  $n = 7$ . Indicate which partitions are conjugates of one another.
- 2.102.** Compute  $p(8, 3)$  by direct enumeration and by using a recursion.
- 2.103.** Use Euler's recursion to compute  $p(k)$  for 13, 14, 15, 16 (see Figure 2.19).
- 2.104.** (a) Write down all the set partitions and rook placements counted by  $S(5, 2)$ . (b) List all the set partitions and equivalence relations counted by  $B(4)$ . (c) Draw all the wrook placements counted by  $s'(4, 2)$ .
- 2.105.** Compute  $S(9, k)$  for  $0 \leq k \leq 9$  and  $S(10, k)$  for  $0 \leq k \leq 10$  (use Figure 2.21).
- 2.106.** Compute the Bell number  $B(k)$  for  $k = 9, 10, 11, 12$  (use Figure 2.21).
- 2.107.** Compute  $s(8, k)$  for  $0 \leq k \leq 8$  (use Figure 2.24).
- 2.108.** (a) Find a combinatorial proof of the formula  $\sum_{i=1}^n i = n(n+1)/2$ . (b) Can you prove  $\sum_{i=1}^n i^2 = n(n+1)(2n+1)/6$  combinatorially?
- 2.109.** Prove the identity  $k \binom{n}{k} = n \binom{n-1}{k-1} = (n-k+1) \binom{n}{k-1}$  algebraically and combinatorially (where  $1 \leq k \leq n$ ).

**2.110.** Suppose  $X$  is an  $n$ -element set. Count the number of: (a) relations on  $X$ ; (b) reflexive relations on  $X$ ; (c) irreflexive relations on  $X$ ; (d) symmetric relations on  $X$ ; (e) irreflexive and symmetric relations on  $X$ ; (f) antisymmetric relations on  $X$ .

**2.111.** Let  $X$  be a nine-element set and  $Y$  a four-element set. (a) How many functions map  $X$  into  $Y$ ? (b) How many functions map  $Y$  into  $X$ ? (c) How many surjections are there from  $X$  onto  $Y$ ? (d) How many injections are there from  $Y$  into  $X$ ?

**2.112.** Verify equations (2.6), (2.7), and (2.8) by direct calculation for  $n = 3$  and  $n = 4$ .

**2.113.** (a) Find the rook placement associated to the set partition

$$\{\{2, 5\}, \{1, 4, 7, 10\}, \{3\}, \{6, 8\}, \{9\}\}$$

by the bijection in 2.64. (b) Find the set partition associated to the following rook placement:

$R$					
			$R$		
	$R$				

**2.114.** Let  $f : \{1, 2, \dots, 7\} \rightarrow \{1, 2, 3\}$  be the surjection given by  $f(1) = 3$ ,  $f(2) = 3$ ,  $f(3) = 1$ ,  $f(4) = 3$ ,  $f(5) = 2$ ,  $f(6) = 3$ ,  $f(7) = 1$ . In the second proof of 2.58, what choice sequence can be used to construct  $f$ ?

**2.115.** How many compositions of 20 only use parts of sizes 1, 3, or 5?

**2.116.** Use the recursion 2.26 for multisets to prove by induction that the number of  $k$ -element multisets using an  $n$ -element alphabet is  $M(n, k) = \frac{(k+n-1)!}{k!(n-1)!}$ .

**2.117.** Given  $a, b, c, n \in \mathbb{N}^+$  with  $a + b + c = n$ , prove combinatorially that  $\binom{n}{a, b, c} = \sum_{k=0}^c \left[ \binom{n-k-1}{a-1, b, c-k} + \binom{n-k-1}{a, b-1, c-k} \right]$ .

**2.118.** Complete the proof of 2.31 by proving  $T_m(a, b) = \frac{b-ma+1}{b+a+1} \binom{a+b+1}{a}$  by induction.

**2.119.** Show that  $|S_n^\tau| = C_n$  for (a)  $\tau = 132$ ; (b)  $\tau = 213$ ; (c)  $\tau = 312$ . (d) Convert the binary tree in Figure 2.12 to a  $\tau$ -avoiding permutation for each of these choices of  $\tau$ .

**2.120.** (a) Let  $G_n$  be the set of lists of integers  $(g_0, g_1, \dots, g_{n-1})$  where  $g_0 = 0$ , each  $g_i \geq 0$ , and  $g_{i+1} \leq g_i + 1$  for all  $i < n - 1$ . Prove that  $|G_n| = C_n$ . (b) For  $m \in \mathbb{N}^+$ , let  $G_n^{(m)}$  be the set of lists of integers  $(g_0, g_1, \dots, g_{n-1})$  where  $g_0 = 0$ , each  $g_i \geq 0$ , and  $g_{i+1} \leq g_i + m$  for all  $i < n - 1$ . Prove that  $|G_n^{(m)}| = T_m(n, mn)$ , the number of lattice paths from  $(0, 0)$  to  $(n, mn)$  that never go below the line  $y = mx$ .

**2.121.** Consider the 231-avoiding permutation  $w = 1\ 5\ 2\ 4\ 3\ 11\ 7\ 6\ 10\ 8\ 9$ . Use recursive bijections based on the Catalan recursion to map  $w$  to objects of the following kinds: (a) a Dyck path; (b) a binary tree; (c) a 312-avoiding permutation (see 2.119); (d) an element of  $G_n$  (see 2.120).

**2.122.** Let  $\pi$  be the Dyck path NNENEENNENNNENNEEENENEEEE. Use recursive bijections based on the Catalan recursion to map  $\pi$  to objects of the following kinds: (a) a binary tree; (b) a 231-avoiding permutation; (c) a 213-avoiding permutation (see 2.119).

**2.123.** Show that the number of possible rhyme schemes for an  $n$ -line poem using  $k$  different rhyme syllables is the Stirling number  $S(n, k)$ . (For example, ABABCD CDEF EFGG is a rhyme scheme with  $n = 14$  and  $k = 7$ .)

**2.124.** (a) Find explicit formulas for  $S(n, k)$  when  $k = 1, 2, n - 1$ , and  $n$ . Prove your formulas using counting arguments. (b) Repeat part (a) for  $\text{Surj}(n, k)$ .

**2.125.** Give a combinatorial proof of the identity  $kS(n, k) = \sum_{j=1}^n \binom{n}{j} S(n-j, k-1)$ , where  $1 \leq k \leq n$ .

**2.126.** Prove  $C_n = \sum_{k \in \mathbb{N}: 0 \leq k \leq n/2} T(k, n-k)^2$  for  $n \geq 1$ .

**2.127.** Consider lattice paths that can take unit steps up (N), down (S), left (W), or right (E), with self-intersections allowed. How many such paths begin and end at  $(0, 0)$  and have 10 steps?

**2.128.** Use 2.52, 2.63, and the ideas in 2.39 to give a recursive definition of a bijection between rook placements counted by  $S'(n, k)$  and set partitions counted by  $S(n, k)$ . Is this bijection the same as the bijection described in 2.64?

**2.129.** Fix  $n \in \mathbb{N}^+$ , let  $\mu$  be an integer partition of length  $\ell(\mu) \leq n$ , and set  $\mu_k = 0$  for  $\ell(\mu) < k \leq n$ . Let  $s'(\mu, k)$  be the number of placements of  $n - k$  non-attacking wrooks on the board  $\text{dg}(\mu)$ . (a) Find a summation formula for  $s'(\mu, k)$  analogous to 2.75. (b) Prove that

$$(x + \mu_1)(x + \mu_2) \cdots (x + \mu_n) = \sum_{k=0}^n s'(\mu, k) x^k.$$

(c) For  $n = 7$  and  $\mu = (8, 5, 3, 3, 1)$ , find  $s'(\mu, k)$  for  $0 \leq k \leq 7$ .

**2.130.** (a) Show that the Fibonacci number  $f_{n-1}$  (see 2.23) is the number of compositions of  $n$  in which every part has size 1 or 2. (b) Show that  $f_n$  is the number of subsets of  $\{1, 2, \dots, n\}$  that do not contain two consecutive integers. (c) Combine (b) with 1.113 to deduce a summation formula for  $f_n$ .

**2.131.** (a) Show that the sequence  $a_n = f_{2n}$  (see 2.23) satisfies the recursion  $a_n = 3a_{n-1} - a_{n-2}$  for  $n \geq 2$ . What are the initial conditions? (b) Show that  $a_n$  is the number of words in  $\{A, B, C\}^n$  in which A is never immediately followed by B.

**2.132.** For  $n \geq 0$ , let  $a_n$  be the number of words in  $\{1, 2, \dots, k\}^n$  in which 1 is never immediately followed by 2. Find a recursion satisfied by the sequence  $(a_n : n \geq 0)$ , and prove it with a suitable bijection.

**2.133.** (a) Give algebraic and combinatorial proofs of the identity

$$x^n - 1 = (x - 1)1 + (x - 1)x + (x - 1)x^2 + \cdots + (x - 1)x^{n-1} \quad (x \in \mathbb{R}).$$

(b) Deduce a formula for  $\sum_{m=0}^{\infty} x^m$ , valid for real numbers  $x$  with  $|x| < 1$ .

**2.134.** Define  $F_0 = 0$ ,  $F_1 = 1$ , and  $F_n = F_{n-1} + F_{n-2}$  for all  $n > 1$  (so  $f_n = F_{n+2}$  for  $n \geq 0$ ). Give algebraic or combinatorial proofs of the following formulas. (a)  $F_n = (\phi^n - \psi^n)/\sqrt{5}$ , where  $\phi = (1 + \sqrt{5})/2$ ,  $\psi = (1 - \sqrt{5})/2$ . (b)  $\sum_{k=0}^n F_k = F_{n+2} - 1$ . (c)  $\sum_{k=0}^{n-1} F_{2k+1} = F_{2n}$ . (d)  $\sum_{k=0}^n F_{2k} = F_{2n+1} - 1$ .

**2.135.** Give a combinatorial proof of equation (2.11) by interpreting both sides as counting a suitable collection of functions.

**2.136.** Let  $C_{n,k}$  be the number of Dyck paths of order  $n$  that end with exactly  $k$  east steps. Prove the recursion

$$C_{n,k} = \sum_{r=1}^{n-k} \binom{k-1+r}{k-1, r} C_{n-k, r}.$$

**2.137.** Let  $p$  be prime. Prove that  $\binom{p}{k}$  is divisible by  $p$  for  $0 < k < p$ . Can you find a combinatorial proof?

**2.138.** *Fermat's Little Theorem* states that  $a^p \equiv a \pmod{p}$  for  $a \in \mathbb{N}^+$  and  $p$  prime. Prove this by expanding  $a^p = (1 + 1 + \cdots + 1)^p$  using the multinomial theorem (cf. 2.137).

**2.139. Ordered Set Partitions.** An *ordered set partition* of a set  $X$  is a sequence  $P = (T_1, T_2, \dots, T_k)$  of distinct sets such that  $\{T_1, T_2, \dots, T_k\}$  is a set partition of  $X$ . Let  $B_o(n)$  be the number of ordered set partitions of an  $n$ -element set. (a) Show  $B_o(n) = \sum_{k=1}^n k! S(n, k)$  for  $n \geq 1$ . (b) Find a recursion relating  $B_o(n)$  to values of  $B_o(m)$  for  $m < n$ . (c) Compute  $B_o(n)$  for  $0 \leq n \leq 5$ .

**2.140.** (a) Let  $B_1(n)$  be the number of set partitions of an  $n$ -element set such that no block of the partition has size 1. Find a recursion and initial conditions for  $B_1(n)$ , and use these to compute  $B_1(n)$  for  $1 \leq n \leq 6$ . (b) Let  $S_1(n, k)$  be the number of set partitions as in (a) with  $k$  blocks. Find a recursion and initial conditions for  $S_1(n, k)$ .

**2.141.** Let  $p_d(n, k)$  be the set of integer partitions of  $n$  with first part  $k$  and all parts *distinct*. Find a recursion and initial conditions for  $p_d(n, k)$ .

**2.142.** Let  $p_o(n, k)$  be the set of integer partitions of  $n$  with first part  $k$  and all parts *odd*. Find a recursion and initial conditions for  $p_o(n, k)$ .

**2.143.** Let  $q(n, k)$  be the number of integer partitions  $\mu$  of length  $k$  and area  $n$  such that  $\mu' = \mu$  (such partitions are called *self-conjugate*). Find a recursion and initial conditions for  $q(n, k)$ .

**2.144.** Verify the statement made in 2.76 that any indexed collection of polynomials  $\{p_n(x) : n \geq 0\}$  such that  $\deg(p_n) = n$  for all  $n$  is a basis for the real vector space of polynomials in one variable with real coefficients.

**2.145.** Verify the following statements about transition matrices from §2.13. (a) If  $B$  and  $C$  are ordered bases of  $W$  and  $A$  is the transition matrix from  $B$  to  $C$ , then  $[v]_C = A[v]_B$  for all  $v \in W$ . (b) If  $A$  is the transition matrix from  $B$  to  $C$ , then  $A$  is invertible, and  $A^{-1}$  is the transition matrix from  $C$  to  $B$ .

**2.146.** Complete the proof of 2.55 by verifying that: (a)  $\phi(P) \in \mathcal{B}$  for all  $P \in \mathcal{A}$ ; (b)  $\phi'(R) \in \mathcal{A}$  for all  $R \in \mathcal{B}$ ; (c)  $\phi \circ \phi' = \text{id}_{\mathcal{B}}$ ; (d)  $\phi' \circ \phi = \text{id}_{\mathcal{A}}$ .

**2.147.** Consider a product  $x_1 \times x_2 \times \cdots \times x_n$  where the binary operation  $\times$  is not necessarily associative. Show that the number of ways to parenthesize this expression is the Catalan number  $C_{n-1}$ . For example, the five possible parenthesizations when  $n = 4$  are

$$\begin{aligned} &(((x_1 \times x_2) \times x_3) \times x_4), ((x_1 \times x_2) \times (x_3 \times x_4)), (x_1 \times ((x_2 \times x_3) \times x_4)), \\ &(x_1 \times (x_2 \times (x_3 \times x_4))), ((x_1 \times (x_2 \times x_3)) \times x_4). \end{aligned}$$

**2.148. Generalized Associative Law.** Let  $\times$  be an associative binary operation on a set  $S$  (so  $(x \times y) \times z = x \times (y \times z)$  for all  $x, y, z \in S$ ). Given  $x_1, \dots, x_n \in S$ , recursively define  $\prod_{i=1}^1 x_i = x_1$  and  $\prod_{i=1}^n x_i = (\prod_{i=1}^{n-1} x_i) \times x_n$ . Prove that every parenthesization of the expression  $x_1 \times x_2 \times \cdots \times x_n$  evaluates to  $\prod_{i=1}^n x_i$  (use strong induction). This result justifies the omission of parentheses in expressions of this kind.

**2.149. Generalized Commutative Law.** Let  $+$  be an associative and commutative binary operation on a set  $S$ . Prove that for any bijection  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  and any elements  $x_1, \dots, x_n \in S$ ,

$$x_1 + x_2 + \cdots + x_n = x_{f(1)} + x_{f(2)} + \cdots + x_{f(n)}.$$

**2.150.** For each positive integer  $n$ , let  $\mathbb{Z}_n = \{0, 1, 2, \dots, n-1\}$ . Define binary operations  $\oplus$  and  $\otimes$  on  $\mathbb{Z}_n$  by letting  $a \oplus b = (a + b) \bmod n$  and  $a \otimes b = (a \cdot b) \bmod n$ , where  $c \bmod n$  denotes the unique remainder in  $\mathbb{Z}_n$  when  $c$  is divided by  $n$ . (a) Prove that  $\mathbb{Z}_n$  with these operations is a commutative ring. (b) Prove that  $\mathbb{Z}_n$  is a field iff  $n$  is a prime number.

**2.151.** Let  $R$  be a ring, and let  $M_n(R)$  be the set of all  $n \times n$  matrices with entries in  $R$ . Define matrix addition and multiplication as follows. Writing  $A(i, j)$  for the  $i, j$ -entry of a matrix  $A$ , let  $(A + B)(i, j) = A(i, j) + B(i, j)$  and  $(AB)(i, j) = \sum_{k=1}^n A(i, k)B(k, j)$ . Verify the ring axioms in 2.2 for  $M_n(R)$ . Show that this ring is non-commutative whenever  $n > 1$  and  $|R| > 1$ . If  $R$  is finite, what is  $|M_n(R)|$ ?

**2.152.** Let  $R$  be a ring, and let  $R[x]$  be the set of all one-variable “formal” polynomials with coefficients in  $R$ . Define polynomial addition and multiplication as follows. If  $p = \sum_{i \geq 0} a_i x^i$  and  $q = \sum_{j \geq 0} b_j x^j$  with  $a_i, b_j \in R$  (and  $a_i = 0, b_j = 0$  for large enough  $i, j$ ), set  $p + q = \sum_i (a_i + b_i) x^i$  and  $pq = \sum_k c_k x^k$ , where  $c_k = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0$ . Also, by definition,  $p = q$  means  $a_i = b_i$  for all  $i \geq 0$ . Verify the ring axioms in 2.2 for  $R[x]$ .

**2.153.** (a) Let  $R$  be the set of all functions  $f : \mathbb{Z} \rightarrow \mathbb{Z}$ . Given  $f, g \in R$ , define  $f \oplus g : \mathbb{Z} \rightarrow \mathbb{Z}$  by  $(f \oplus g)(n) = f(n) + g(n)$  for all  $n \in \mathbb{Z}$ , and define  $f \circ g$  by  $(f \circ g)(n) = f(g(n))$  (composition of functions). Show that  $(R, \oplus, \circ)$  satisfies all the ring axioms in 2.2 except commutativity of multiplication and the left distributive law. (b) Let  $S$  be the set of  $f \in R$  such that  $f(m+n) = f(m) + f(n)$  for all  $m, n \in \mathbb{Z}$ . Prove that  $(S, \oplus, \circ)$  is a ring.

**2.154.** Prove the binomial theorem 2.14 by induction on  $n$ . Mark each place in your proof where you use the hypothesis  $xy = yx$ .

**2.155.** Prove the commutative multinomial theorem 2.12 using the binomial theorem and induction on  $s$ .

**2.156.** Let  $f, g$  be smooth functions of  $x$  (which means  $f$  and  $g$  have derivatives of all orders). Recall the product rule:  $D(fg) = D(f)g + fD(g)$ , where  $D$  denotes differentiation with respect to  $x$ . (a) Prove that the  $n$ th derivative of  $fg$  is given by

$$D^n(fg) = \sum_{k=0}^n \binom{n}{k} D^k(f) D^{n-k}(g).$$

(b) Find and prove a similar formula for  $D^n(f_1 f_2 \cdots f_s)$ , where  $f_1, \dots, f_s$  are smooth functions.

**2.157.** (a) Given a field  $F$ , an element  $c \in F$ , and a polynomial  $p \in F[x]$ , show that  $p(c) = 0$  iff  $x - c$  divides  $p$  in  $F[x]$ . (b) Show that if  $p \in F[x]$  is a nonzero polynomial with more than  $N$  roots in  $F$ , then  $\deg(p) > N$ . (c) Show that (b) can fail if  $F$  is a commutative ring that is not a field.

**2.158.** Let  $p, q \in \mathbb{R}[x_1, \dots, x_n]$  be multivariable polynomials such that  $p(m_1, \dots, m_n) = q(m_1, \dots, m_n)$  for all  $m_1, \dots, m_n \in \mathbb{N}^+$ . Prove that  $p = q$ .

**2.159.** (a) Give a combinatorial proof of the multinomial theorem 2.12, assuming that  $z_1, z_2, \dots, z_s$  are positive integers. (b) Deduce that this theorem is also valid for all  $z_1, \dots, z_s \in \mathbb{R}$ .

**2.160.** Let  $A_n$  be the set of lattice paths from  $(0,0)$  to  $(n,n)$  that take exactly one north step below the line  $y = x$ . What is  $|A_n|$ ?

**2.161.** Prove: for  $n \in \mathbb{N}$ ,  $\sum_{k=0}^n \binom{2k}{k} \binom{2n-2k}{n-k} = 4^n$ .

---

## Notes

The book by Gould [58] contains an extensive, systematic list of binomial coefficient identities. More recently, Petkovsek, Wilf and Zeilberger [104] developed an algorithm, called the WZ-method, that can automatically evaluate many hypergeometric summations (which include binomial coefficient identities) or prove that such a summation has no closed form. For more information on hypergeometric series, see Koepf [80].

A wealth of information about integer partitions, including a discussion of the Hardy-Rademacher-Ramanujan formula 2.49, may be found in [5]. There is a vast literature on pattern-avoiding permutations; for more information on this topic, consult Bona [15].

A great many combinatorial interpretations have been discovered for the Catalan numbers  $C_n$ . A partial list appears in Exercise 6.19 of Stanley [127, Vol. 2]; this list continues in the “Catalan addendum,” which currently resides on the Internet at the following location:

<http://www-math.mit.edu/~rstan/ec/catadd.pdf>

This page intentionally left blank

---

## Counting Problems in Graph Theory

---

Graph theory is a branch of discrete mathematics that studies networks composed of a number of sites (vertices) linked together by connecting arcs (edges). This chapter studies some enumeration problems that arise in graph theory. We begin by defining fundamental graph-theoretic concepts such as walks, paths, cycles, vertex degrees, connectivity, forests, and trees. This will lead to a discussion of various enumeration problems involving different kinds of trees. Aided by ideas from matrix theory, we will count walks in a graph, spanning trees of a graph, and Eulerian tours. We also investigate the chromatic polynomial of a graph, which counts the number of ways of coloring the vertices such that no two vertices joined by an edge receive the same color.

---

### 3.1 Graphs and Digraphs

Intuitively, a graph is a mathematical model for a network consisting of a collection of nodes and connections that link certain pairs of nodes. For example, the nodes could be cities and the connections could be roads between cities. For another example, the nodes could be computers and the connections could be network links between computers. For a third example, the nodes could be species in an ecosystem and the connections could be predator-prey relationships between species. Or the nodes could be tasks and the connections could be dependencies among the tasks. There is an unlimited variety of such applications that lead naturally to graph models. We now give the formal mathematical definitions underlying such models.

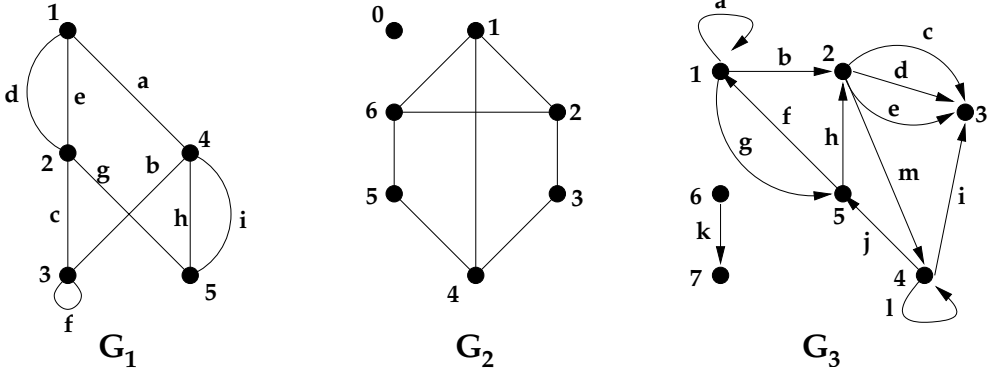
**3.1. Definition: Graphs.** A *graph* is an ordered triple  $G = (V, E, \epsilon)$ , where:  $V = V(G)$  is a finite, nonempty set called the *vertex set* of  $G$ ;  $E = E(G)$  is a finite set called the *edge set* of  $G$ ; and  $\epsilon : E \rightarrow \mathcal{P}(V)$  is a function called the *endpoint function* such that, for all  $e \in E$ ,  $\epsilon(e)$  is either a one-element subset of  $V$  or a two-element subset of  $V$ . If  $\epsilon(e) = \{v\}$ , we call the edge  $e$  a *loop at vertex  $v$* . If  $\epsilon(e) = \{v, w\}$ , we call  $v$  and  $w$  the *endpoints of  $e$*  and say that  $e$  is an edge *from  $v$  to  $w$* . We also say that  $v$  and  $w$  are *adjacent in  $G$* ,  $v$  and  $w$  are *joined by  $e$* , and  $e$  is *incident to  $v$  and  $w$* .

We visualize a graph  $G = (V, E, \epsilon)$  by drawing a collection of dots labeled by the elements  $v \in V$ . For each edge  $e \in E$  with  $\epsilon(e) = \{v, w\}$ , we draw a line or curved arc labeled  $e$  between the two dots labeled  $v$  and  $w$ . Similarly, if  $\epsilon(e) = \{v\}$ , we draw a loop labeled  $e$  based at the dot labeled  $v$ .

**3.2. Example.** The left-hand drawing in Figure 3.1 represents the graph defined formally by the ordered triple

$$G_1 = (\{1, 2, 3, 4, 5\}, \{a, b, c, d, e, f, g, h, i\}, \epsilon),$$



**FIGURE 3.1**

A graph, a simple graph, and a digraph.

where  $\epsilon$  acts as follows:

$$\begin{aligned} a &\mapsto \{1, 4\}, & b &\mapsto \{4, 3\}, & c &\mapsto \{2, 3\}, & d &\mapsto \{1, 2\}, & e &\mapsto \{1, 2\}, \\ f &\mapsto \{3\}, & g &\mapsto \{2, 5\}, & h &\mapsto \{4, 5\}, & i &\mapsto \{4, 5\}. \end{aligned}$$

Edge  $f$  is a loop at vertex 3; edges  $h$  and  $i$  both go between vertices 4 and 5; vertices 1 and 4 are adjacent, but vertices 2 and 4 are not.

In many applications, there are no loop edges, and there is never more than one edge between the same two vertices. This means that the endpoint function  $\epsilon$  is a one-to-one map into the set of two-element subsets of  $V$ . So we can identify each edge  $e$  with its set of endpoints  $\epsilon(e)$ . This leads to the following simplified model in which edges are not explicitly named and there is no explicit endpoint function.

**3.3. Definition: Simple Graphs.** A *simple graph* is a pair  $G = (V, E)$ , where  $V$  is a finite nonempty set and  $E$  is a set of two-element subsets of  $V$ . We continue to use all the terminology introduced in 3.1.

**3.4. Example.** The central drawing in Figure 3.1 depicts the simple graph  $G_2$  with vertex set  $V(G_2) = \{0, 1, 2, 3, 4, 5, 6\}$  and edge set

$$E(G_2) = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{5, 6\}, \{1, 6\}, \{2, 6\}, \{1, 4\}\}.$$

To model certain situations (such as predator-prey relationships, or one-way streets in a city), we need to introduce a *direction* on each edge. This leads to the notion of a digraph (directed graph).

**3.5. Definition: Digraphs.** A *digraph* is an ordered triple  $D = (V, E, \epsilon)$ , where  $V$  is a finite nonempty set of *vertices*,  $E$  is a finite set of *edges*, and  $\epsilon : E \rightarrow V \times V$  is the *endpoint function*. If  $\epsilon(e)$  is the *ordered pair*  $(v, w)$ , we say that  $e$  is an edge *from*  $v$  *to*  $w$ .

In a digraph, an edge from  $v$  to  $w$  is not an edge from  $w$  to  $v$  when  $v \neq w$ , since  $(v, w) \neq (w, v)$ . On the other hand, in a graph, an edge from  $v$  to  $w$  is also an edge from  $w$  to  $v$ , since  $\{v, w\} = \{w, v\}$ .

**3.6. Example.** The right-hand drawing in Figure 3.1 displays a typical digraph. In this digraph,  $\epsilon(j) = (4, 5)$ ,  $\epsilon(a) = (1, 1)$ , and so on. There are three edges from 2 to 3, but no edges from 3 to 2. There are edges in both directions between vertices 1 and 5.

As before, we can eliminate specific reference to the endpoint function of a digraph if there are no multiple edges with the same starting vertex and ending vertex.

**3.7. Definition: Simple Digraphs.** A *simple digraph* is an ordered pair  $D = (V, E)$ , where  $V$  is a finite, nonempty set and  $E$  is a subset of  $V \times V$ . Each ordered pair  $(v, w) \in E$  represents an edge in  $D$  from  $v$  to  $w$ . Note that we do allow loops ( $v = w$ ) in a simple digraph.

When investigating structural properties of graphs, the names of the vertices and edges are often irrelevant. The concept of graph isomorphism lets us identify graphs that are “the same” except for the names used for the vertices and edges.

**3.8. Definition: Graph Isomorphism.** Given two graphs  $G = (V, E, \epsilon)$  and  $H = (W, F, \eta)$ , a *graph isomorphism* from  $G$  to  $H$  consists of two bijections  $f : V \rightarrow W$  and  $g : E \rightarrow F$  such that, for all  $e \in E$ , if  $\epsilon(e) = \{v, w\}$  then  $\eta(g(e)) = \{f(v), f(w)\}$  (we allow  $v = w$  here). Digraph isomorphisms are defined similarly:  $\epsilon(e) = (v, w)$  implies  $\eta(g(e)) = (f(v), f(w))$ . We say  $G$  and  $H$  are *isomorphic*, written  $G \cong H$ , iff there exists a graph isomorphism from  $G$  to  $H$ .

In the case of *simple* graphs  $G = (V, E)$  and  $H = (W, F)$ , a graph isomorphism can be viewed as a bijection  $f : V \rightarrow W$  that induces a bijection between the edge sets  $E$  and  $F$ . More specifically, this means that for all  $v, w \in V$ ,  $\{v, w\} \in E$  iff  $\{f(v), f(w)\} \in F$ .

## 3.2 Walks and Matrices

We can travel through a graph by following a succession of edges. Formalizing this idea leads to the concept of a walk.

**3.9. Definition: Walks, Paths, Cycles.** Let  $G = (V, E, \epsilon)$  be a graph or digraph. A *walk* in  $G$  is a sequence

$$W = (v_0, e_1, v_1, e_2, v_2, \dots, e_s, v_s)$$

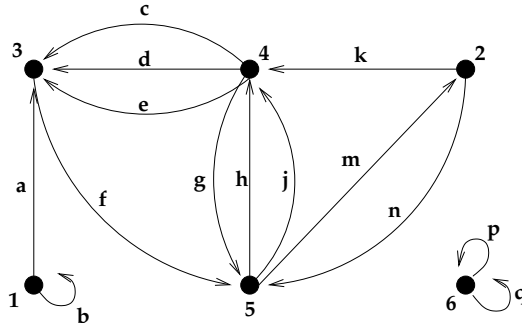
where  $s \geq 0$ ,  $v_i \in V$  for all  $i$ ,  $e_i \in E$  for all  $i$ , and  $e_i$  is an edge from  $v_{i-1}$  to  $v_i$  for  $1 \leq i \leq s$ . We say that  $W$  is a *walk of length  $s$  from  $v_0$  to  $v_s$* . The walk  $W$  is *closed* iff  $v_0 = v_s$ . The walk  $W$  is a *path* iff the vertices  $v_0, v_1, \dots, v_s$  are pairwise distinct (which forces the edges  $e_i$  to be distinct as well). The walk  $W$  is a *cycle* iff  $s > 0$ ,  $v_1, \dots, v_s$  are distinct,  $e_1, \dots, e_s$  are distinct, and  $v_0 = v_s$ . A *k-cycle* is a cycle of length  $k$ . In the case of simple graphs and simple digraphs, the edges  $e_i$  are determined uniquely by their endpoints. So, in the simple case, we can regard a walk as a sequence of vertices  $(v_0, v_1, \dots, v_s)$  such that there is an edge from  $v_{i-1}$  to  $v_i$  in  $G$  for  $1 \leq i \leq s$ .

**3.10. Remark.** When considering cycles in a digraph, we usually *identify* two cycles that are cyclic shifts of one another (unless we need to keep track of the starting vertex of the cycle). Similarly, we identify cycles in a graph that are cyclic shifts or reversals of one another.

**3.11. Example.** In the graph  $G_1$  from Figure 3.1,

$$W_1 = (2, c, 3, f, 3, b, 4, h, 5, i, 4, i, 5)$$

is a walk of length 6 from vertex 2 to vertex 5. In the simple graph  $G_2$  in the same figure,

**FIGURE 3.2**

Digraph used to illustrate adjacency matrices.

$W_2 = (1, 6, 2, 3, 4, 5)$  is a walk and a path of length 5, whereas  $C = (6, 5, 4, 3, 2, 6)$  is a 5-cycle. We usually identify  $C$  with the cycles  $(5, 4, 3, 2, 6, 5)$ ,  $(6, 2, 3, 4, 5, 6)$ , etc. In the digraph  $G_3$ ,

$$W_3 = (1, a, 1, g, 5, h, 2, m, 4, j, 5, h, 2, d, 3)$$

is a walk from vertex 1 to vertex 3;  $(5, h, 2, m, 4, j, 5)$  is a 3-cycle;  $(4, l, 4)$  is a 1-cycle; and  $(5, f, 1, g, 5)$  is a 2-cycle. Observe that 1-cycles are the same as loop edges, and 2-cycles cannot exist in simple graphs or simple digraphs. For any vertex  $v$  in a graph or digraph,  $(v)$  is a walk of length zero from  $v$  to  $v$ , which is a path but not a cycle.

We can now formulate our first counting problem: how many walks of a given length are there between two given vertices in a graph or digraph? We will develop an algebraic solution to this problem in which concatenation of walks is modeled by multiplication of suitable matrices.

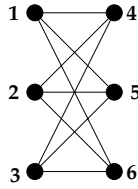
**3.12. Definition: Adjacency Matrix.** Let  $G$  be a graph or digraph with vertex set  $X = \{x_i : 1 \leq i \leq n\}$ . The *adjacency matrix* of  $G$  (relative to the given indexing of the vertices) is the  $n \times n$  matrix  $A$  whose  $i, j$ -entry  $A(i, j)$  is the number of edges in  $G$  from  $x_i$  to  $x_j$ .

**3.13. Example.** The adjacency matrix for the digraph  $G$  in Figure 3.2 is

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

**3.14. Example.** If  $G$  is a graph, edges from  $v$  to  $w$  are the same as edges from  $w$  to  $v$ . So, the adjacency matrix for  $G$  is a *symmetric* matrix ( $A(i, j) = A(j, i)$  for all  $i, j$ ). If  $G$  is a simple graph, the adjacency matrix consists of all 1's and 0's with zeroes on the main diagonal. For example, the adjacency matrix of the simple graph in Figure 3.3 is

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}.$$


**FIGURE 3.3**

A simple graph.

Recall from linear algebra the definition of the product of two matrices.

**3.15. Definition: Matrix Multiplication.** Suppose  $A$  is an  $m \times n$  matrix and  $B$  is an  $n \times p$  matrix. Then  $AB$  is the  $m \times p$  matrix whose  $i, j$ -entry is

$$(AB)(i, j) = \sum_{k=1}^n A(i, k)B(k, j) \quad (1 \leq i \leq m, 1 \leq j \leq p). \quad (3.1)$$

Matrix multiplication is associative, so we can write a product of three or more (compatible) matrices without any parentheses. The next theorem gives a formula for the general entry in such a product.

**3.16. Theorem: Product of Several Matrices.** Assume  $A_1, \dots, A_s$  are matrices such that  $A_i$  has dimensions  $n_{i-1} \times n_i$ . Then  $A_1 A_2 \cdots A_s$  is the  $n_0 \times n_s$  matrix whose  $k_0, k_s$ -entry is

$$(A_1 A_2 \cdots A_s)(k_0, k_s) = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \cdots \sum_{k_{s-1}=1}^{n_{s-1}} A_1(k_0, k_1) A_2(k_1, k_2) A_3(k_2, k_3) \cdots A_s(k_{s-1}, k_s) \quad (3.2)$$

for all  $1 \leq k_0 \leq n_0, 1 \leq k_s \leq n_s$ .

*Proof.* We use induction on  $s$ . The case  $s = 1$  is immediate, and the case  $s = 2$  is the definition of matrix multiplication (after a change in notation). Assume  $s > 2$  and that (3.2) is known to hold for the product  $B = A_1 A_2 \cdots A_{s-1}$ . We can think of the given product  $A_1 A_2 \cdots A_{s-1} A_s$  as the binary product  $BA_s$ . Therefore, using (3.1), the  $k_0, k_s$ -entry of  $A_1 A_2 \cdots A_s$  is

$$\begin{aligned} (BA_s)(k_0, k_s) &= \sum_{k=1}^{n_{s-1}} B(k_0, k) A_s(k, k_s) \\ &= \sum_{k=1}^{n_{s-1}} \left( \sum_{k_1=1}^{n_1} \cdots \sum_{k_{s-2}=1}^{n_{s-2}} A_1(k_0, k_1) A_2(k_1, k_2) \cdots A_{s-1}(k_{s-2}, k) \right) A_s(k, k_s) \\ &= \sum_{k_1=1}^{n_1} \cdots \sum_{k_{s-2}=1}^{n_{s-2}} \sum_{k=1}^{n_{s-1}} A_1(k_0, k_1) A_2(k_1, k_2) \cdots A_{s-1}(k_{s-2}, k) A_s(k, k_s). \end{aligned}$$

Replacing  $k$  by  $k_{s-1}$  in the innermost summation, we obtain the result.  $\square$

Taking all  $A_i$ 's in the theorem to be the same matrix  $A$ , we obtain the following formula for the entries of the powers of a given square matrix.

**3.17. Corollary: Powers of a Matrix.** Suppose  $A$  is an  $n \times n$  matrix. For each integer  $s > 0$ ,

$$A^s(i, j) = \sum_{k_1=1}^n \cdots \sum_{k_{s-1}=1}^n A(i, k_1)A(k_1, k_2) \cdots A(k_{s-2}, k_{s-1})A(k_{s-1}, j) \quad (1 \leq i, j \leq n). \quad (3.3)$$

The preceding formula may appear unwieldy. But it is precisely the tool we need to count walks in graphs.

**3.18. Theorem: Enumeration of Walks.** Let  $G$  be a graph or digraph with vertex set  $X = \{x_1, \dots, x_n\}$ , and let  $A$  be the adjacency matrix of  $G$ . For all  $i, j \leq n$  and all  $s \geq 0$ , the  $i, j$ -entry of  $A^s$  is the number of walks of length  $s$  in  $G$  from  $x_i$  to  $x_j$ .

*Proof.* The result holds for  $s = 0$ , since  $A^0 = I_n$  (the  $n \times n$  identity matrix) and there is exactly one walk of length zero from any vertex to itself. Now suppose  $s > 0$ . A walk of length  $s$  from  $x_i$  to  $x_j$  will visit  $s - 1$  intermediate vertices (not necessarily distinct from each other or from  $x_i$  or  $x_j$ ). Let  $(x_i, x_{k_1}, \dots, x_{k_{s-1}}, x_j)$  be the ordered list of vertices visited by the walk. To build such a walk, we choose any edge from  $x_i$  to  $x_{k_1}$  in  $A(i, k_1)$  ways; then we choose any edge from  $x_{k_1}$  to  $x_{k_2}$  in  $A(k_1, k_2)$  ways; and so on. By the product rule, the total number of walks associated to this vertex sequence is  $A(i, k_1)A(k_1, k_2) \cdots A(k_{s-1}, j)$ . This formula holds even if there are no walks with this vertex sequence, since some term in the product will be zero in this case. Applying the sum rule produces the right side of (3.3), and the result follows.  $\square$

**3.19. Example.** Consider again the adjacency matrix  $A$  of the digraph  $G$  in Figure 3.2. Some matrix computations show that

$$A^2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 3 & 2 & 1 & 0 \\ 0 & 1 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 2 & 3 & 0 \\ 0 & 0 & 6 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 1 & 1 & 1 & 2 & 1 & 0 \\ 0 & 1 & 6 & 3 & 6 & 0 \\ 0 & 0 & 6 & 1 & 3 & 0 \\ 0 & 3 & 6 & 7 & 3 & 0 \\ 0 & 3 & 3 & 6 & 7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 \end{bmatrix}.$$

So, for example, there are six walks of length 2 from vertex 5 to vertex 3, and there are seven walks of length 3 that start and end at vertex 4.

### 3.3 DAGs and Nilpotent Matrices

Next we consider the question of counting *all* walks (of any length) between two vertices in a digraph. The question is uninteresting for graphs, since the number of walks between two distinct vertices  $v, w$  in a graph is either zero or infinity. This follows since a walk is allowed to repeatedly traverse a particular edge along a path from  $v$  to  $w$ , which leads to arbitrarily long walks from  $v$  to  $w$ . Similarly, if  $G$  is a digraph that contains a cycle, we obtain arbitrarily long walks between two vertices on the cycle by going around the cycle again and again. To rule out these possibilities, we restrict attention to the following class of digraphs.

**3.20. Definition: DAGs.** A *DAG* is a digraph with no cycles. (The acronym DAG stands for “directed acyclic graph.”)

To characterize adjacency matrices of DAGs, we need another concept from matrix theory.

**3.21. Definition: Nilpotent Matrices.** An  $n \times n$  matrix  $A$  is called *nilpotent* iff  $A^s = 0$  for some integer  $s \geq 1$ . The least such integer  $s$  is called the *index of nilpotence* of  $A$ .

Note that if  $A^s = 0$  and  $t \geq s$ , then  $A^t = 0$  also.

**3.22. Example.** The zero matrix is the unique  $n \times n$  matrix with index of nilpotence 1. The square of the nonzero matrix  $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$  is zero, so  $A$  is nilpotent of index 2. Similarly, any matrix

$$B = \begin{bmatrix} 0 & x & y \\ 0 & 0 & z \\ 0 & 0 & 0 \end{bmatrix} \quad (x, y, z \in \mathbb{R})$$

satisfies

$$B^2 = \begin{bmatrix} 0 & 0 & xz \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B^3 = 0,$$

so we obtain examples of matrices that are nilpotent of index 3. The next result generalizes this example.

**3.23. Theorem: Nilpotence of Strictly Triangular Matrices.** Suppose  $A$  is an  $n \times n$  *strictly upper-triangular* matrix, which means  $A(i, j) = 0$  for all  $i \geq j$ . Then  $A$  is nilpotent of index at most  $n$ . A similar result holds for strictly lower-triangular matrices.

*Proof.* It suffices to show that  $A^n$  is the zero matrix. By (3.3), we have

$$A^n(k_0, k_n) = \sum_{k_1=1}^n \cdots \sum_{k_{n-1}=1}^n A(k_0, k_1)A(k_1, k_2) \cdots A(k_{n-1}, k_n)$$

for all  $k_0, k_n \leq n$ . We claim that each term in this sum is zero. Otherwise, there would exist  $k_1, \dots, k_{n-1}$  such that  $A(k_{t-1}, k_t) \neq 0$  for  $1 \leq t \leq n$ . But since  $A$  is strictly upper-triangular, we would then have

$$k_0 < k_1 < k_2 < \cdots < k_n.$$

This cannot occur, since all the  $k_t$ 's are integers between 1 and  $n$ . □

The next theorem reveals the connection between nilpotent matrices and DAGs.

**3.24. Theorem: DAGs and Nilpotent Matrices.** Let  $G$  be a digraph with vertex set  $X = \{x_1, \dots, x_n\}$  and adjacency matrix  $A$ .  $G$  is a DAG iff  $A$  is nilpotent. When  $G$  is a DAG, there exists an ordering of the vertex set  $X$  that makes  $A$  a strictly lower-triangular matrix.

*Proof.* Assume first that  $G$  is not a DAG, so that  $G$  has at least one cycle. Let  $x_i$  be any fixed vertex involved in this cycle, and let  $c \geq 1$  be the length of this cycle. By going around the cycle zero or more times, we obtain walks from  $x_i$  to  $x_i$  of lengths  $0, c, 2c, 3c, \dots$ . By 3.18, it follows that the  $(i, i)$ -entry of  $A^{kc}$  is at least 1, for all  $k \geq 0$ . This fact prevents any positive power of  $A$  from being the zero matrix, so  $A$  is not nilpotent.

Conversely, assume that  $A$  is not nilpotent. Then, in particular,  $A^n \neq 0$ , so there exist indices  $k_0, k_n$  with  $A^n(k_0, k_n) \neq 0$ . Using 3.18 again, we deduce that there is a walk in  $G$  visiting a sequence of vertices

$$(x_{k_0}, x_{k_1}, x_{k_2}, \dots, x_{k_{n-1}}, x_{k_n}).$$

Since  $G$  has only  $n$  vertices, not all of the  $n + 1$  vertices just listed are distinct. If we choose  $i$  minimal and then  $j > i$  minimal such that  $x_{k_i} = x_{k_j}$ , then there is a cycle in  $G$  visiting the vertices  $(x_{k_i}, x_{k_{i+1}}, \dots, x_{k_j})$ . So  $G$  is not a DAG.

We prove the statement about lower-triangular matrices by induction on  $n$ . A one-vertex DAG must have adjacency matrix  $(0)$ , so the result holds for  $n = 1$ . Suppose  $n > 1$  and the result is known for DAGs with  $n - 1$  vertices. Create a walk  $(v_0, e_1, v_1, e_2, v_2, \dots)$  in  $G$  by starting at any vertex and repeatedly following any edge leading away from the current vertex. Since  $G$  has no cycles, the vertices  $v_i$  reached by this walk are pairwise distinct. Since there are only  $n$  available vertices, our walk must terminate at a vertex  $v_j$  with no outgoing edges. Let  $x'_1 = v_j$ . Deleting  $x'_1$  and all edges leading into this vertex will produce an  $(n - 1)$ -vertex digraph  $G'$  that is also a DAG, as one immediately verifies. By induction, there is an ordering  $x'_2, \dots, x'_n$  of the vertices of  $G'$  such that the associated adjacency matrix  $A'$  is strictly lower-triangular. Now, relative to the ordering  $x'_1, x'_2, \dots, x'_n$  of the vertices of  $G$ , the adjacency matrix of  $G$  has the form

$$\left[ \begin{array}{c|ccc} 0 & 0 & \cdots & 0 \\ \hline * & & & \\ \vdots & & A' & \\ * & & & \end{array} \right],$$

and this matrix is strictly lower-triangular.  $\square$

The next result will allow us to count walks of any length in a DAG.

**3.25. Theorem: Inverse of  $I - A$  for Nilpotent  $A$ .** Suppose  $A$  is a nilpotent  $n \times n$  matrix with  $A^s = 0$ . Let  $I$  be the  $n \times n$  identity matrix. Then  $I - A$  is an invertible matrix with inverse

$$(I - A)^{-1} = I + A + A^2 + A^3 + \cdots + A^{s-1}.$$

*Proof.* Let  $B = I + A + A^2 + \cdots + A^{s-1}$ . By the distributive law,

$$(I - A)B = IB - AB = (I + A + A^2 + \cdots + A^{s-1}) - (A + A^2 + \cdots + A^{s-1} + A^s) = I - A^s.$$

Since  $A^s = 0$ , we see that  $(I - A)B = I$ . A similar calculation shows that  $B(I - A) = I$ . Therefore  $B$  is the two-sided matrix inverse of  $I - A$ .  $\square$

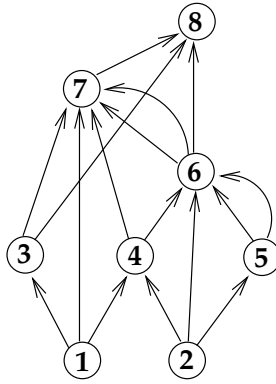
**3.26. Remark.** The previous result for matrices can be remembered by noting the analogy to the geometric series formula for real numbers:

$$(1 - r)^{-1} = \frac{1}{1 - r} = 1 + r + r^2 + \cdots + r^{s-1} + r^s + \cdots \quad (|r| < 1).$$

**3.27. Theorem: Counting Paths in a DAG.** Let  $G$  be a DAG with vertex set  $\{x_1, \dots, x_n\}$  and adjacency matrix  $A$ . For all  $i, j \leq n$ , the total number of paths from  $x_i$  to  $x_j$  in  $G$  (of any length) is the  $i, j$ -entry of  $(I - A)^{-1}$ .

*Proof.* By 3.18, the number of walks of length  $t \geq 0$  from  $x_i$  to  $x_j$  is  $A^t(i, j)$ . Because  $G$  is a DAG, we have  $A^t = 0$  for all  $t \geq n$ . By the sum rule, the total number of walks from  $x_i$  to  $x_j$  is  $\sum_{t=0}^{n-1} A^t(i, j)$ . By 3.25, this number is precisely the  $i, j$ -entry of  $(I - A)^{-1}$ . Finally, one readily confirms that every walk in a DAG must actually be a path.  $\square$

**3.28. Example.** Consider the DAG shown in Figure 3.4. Its adjacency matrix is



**FIGURE 3.4**  
Example of a DAG.

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Using a computer algebra system, we compute

$$(I - A)^{-1} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 5 & 7 \\ 0 & 1 & 0 & 1 & 1 & 4 & 9 & 13 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 0 & 1 & 3 & 4 \\ 0 & 0 & 0 & 0 & 1 & 2 & 4 & 6 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

So, for example, there are 13 paths from vertex 2 to vertex 8, and 4 paths from vertex 5 to vertex 7.

### 3.4 Vertex Degrees

In many situations, one needs to know how many edges lead into or out of each vertex in a digraph.

**3.29. Definition: Indegree and Outdegree.** Let  $G = (V, E, \epsilon)$  be a digraph. For each  $v \in V$ , the *outdegree* of  $v$ , denoted  $\text{outdeg}_G(v)$ , is the number of edges  $e \in E$  from  $v$ ; the *indegree* of  $v$ , denoted  $\text{indeg}_G(v)$ , is the number of edges  $e \in E$  to  $v$ . Formally,

$$\text{outdeg}_G(v) = \sum_{w \in V} \sum_{e \in E} \chi(\epsilon(e) = (v, w)); \quad \text{indeg}_G(v) = \sum_{w \in V} \sum_{e \in E} \chi(\epsilon(e) = (w, v)).$$



A *source* is a vertex of indegree zero. A *sink* is a vertex of outdegree zero.

**3.30. Example.** Let  $G_3$  be the digraph on the right in Figure 3.1. We have

$$\begin{aligned} (\text{indeg}_{G_3}(1), \dots, \text{indeg}_{G_3}(7)) &= (2, 2, 4, 2, 2, 0, 1); \\ (\text{outdeg}_{G_3}(1), \dots, \text{outdeg}_{G_3}(7)) &= (3, 4, 0, 3, 2, 1, 0). \end{aligned}$$

Vertex 6 is the only source, whereas vertices 3 and 7 are sinks. A loop edge at  $v$  contributes 1 to both the indegree and outdegree of  $v$ . The sum of all the indegrees is 13, which is also the sum of all the outdegrees, and is also the number of edges in the digraph. This phenomenon is explained in the next theorem.

**3.31. Theorem: Degree-Sum Formula for Digraphs.** In any digraph  $G = (V, E, \epsilon)$ ,

$$\sum_{v \in V} \text{indeg}_G(v) = |E| = \sum_{v \in V} \text{outdeg}_G(v).$$

*Proof.* By the formal definition of indegree, we have

$$\begin{aligned} \sum_{v \in V} \text{indeg}_G(v) &= \sum_{v \in V} \sum_{w \in V} \sum_{e \in E} \chi(\epsilon(e) = (w, v)) \\ &= \sum_{e \in E} \left( \sum_{w \in V} \sum_{v \in V} \chi(\epsilon(e) = (w, v)) \right) \end{aligned}$$

For each  $e \in E$ , the term in brackets is equal to one for exactly one ordered pair  $(w, v)$ , and is zero otherwise. So the sum evaluates to  $\sum_{e \in E} 1 = |E|$ . The formula involving outdegree is proved similarly.  $\square$

Next we give analogous definitions and results for graphs.

**3.32. Definition: Degree.** Let  $G = (V, E, \epsilon)$  be a graph. For each  $v \in V$ , the *degree of  $v$  in  $G$* , denoted  $\deg_G(v)$ , is the number of edges in  $E$  with  $v$  as an endpoint, where each loop edge at  $v$  is counted twice. Formally,

$$\deg_G(v) = \sum_{e \in E} 2\chi(\epsilon(e) = \{v\}) + \sum_{\substack{w \in V \\ w \neq v}} \sum_{e \in E} \chi(\epsilon(e) = \{v, w\}).$$

The *degree sequence of  $G$* , denoted  $\deg(G)$ , is the multiset  $[\deg_G(v) : v \in V]$ .  $G$  is called  *$k$ -regular* iff every vertex in  $G$  has degree  $k$ .  $G$  is *regular* iff  $G$  is  $k$ -regular for some  $k \geq 0$ .

**3.33. Example.** For the graph  $G_1$  in Figure 3.1, we have

$$(\deg_G(1), \dots, \deg_G(5)) = (3, 4, 4, 4, 3); \quad \deg(G) = [4, 4, 4, 3, 3].$$

The graph in Figure 3.3 is 3-regular. In both of these graphs, the sum of all vertex degrees is 18, which is twice the number of edges in the graph.

**3.34. Theorem: Degree-Sum Formula for Graphs.** For any graph  $G = (V, E, \epsilon)$ ,

$$\sum_{v \in V} \deg_G(v) = 2|E|.$$

*Proof.* First assume  $G$  has no loop edges. Let  $X$  be the set of pairs  $(v, e)$  such that  $v \in V$ ,  $e \in E$ , and  $v$  is an endpoint of  $e$ . On one hand,

$$|X| = \sum_{v \in V} \sum_{e \in E} \chi((v, e) \in X) = \sum_{v \in V} \deg_G(v).$$

On the other hand,

$$|X| = \sum_{e \in E} \sum_{v \in V} \chi((v, e) \in X) = \sum_{e \in E} 2 = 2|E|$$

since every edge has two distinct endpoints. So the result holds in this case.

Next, if  $G$  has  $k$  loop edges, let  $G'$  be  $G$  with these loops deleted. Then

$$\sum_{v \in V} \deg_G(v) = \sum_{v \in V} \deg_{G'}(v) + 2k,$$

since each loop edge increases some vertex degree in the sum by 2. Using the result for loopless graphs,

$$\sum_{v \in V} \deg_G(v) = 2|E(G')| + 2k = 2|E(G)|,$$

since  $G$  has  $k$  more edges than  $G'$ . □

Vertices of low degree are given special names.

**3.35. Definition: Isolated Vertices.** An *isolated vertex* in a graph is a vertex of degree zero.

**3.36. Definition: Leaves.** A *leaf* is a vertex of degree one.

The following result will be used later in our analysis of trees.

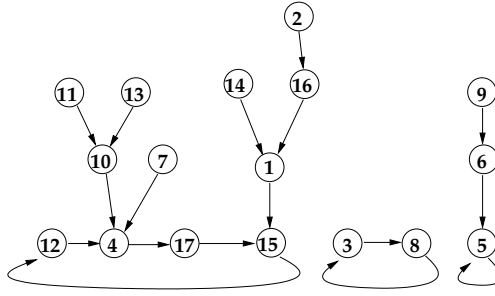
**3.37. Two-Leaf Lemma.** Suppose  $G$  is a graph. One of the following three alternatives must occur: (i)  $G$  has a cycle; (ii)  $G$  has no edges; or (iii)  $G$  has at least two leaves.

*Proof.* Suppose that  $G$  has no cycles and  $G$  has at least one edge; we prove that  $G$  has two leaves. Since  $G$  has no cycles, we can assume  $G$  is simple. Let  $P = (v_0, v_1, \dots, v_s)$  be a path of maximum length in  $G$ . Such a path exists, since  $G$  has only finitely many vertices and edges. Observe that  $s > 0$  since  $G$  has an edge, and  $v_0 \neq v_s$ . Note that  $\deg(v_s) \geq 1$  since  $s > 0$ . Assume  $v_s$  is not a leaf. Then there exists a vertex  $w \neq v_{s-1}$  that is adjacent to  $v_s$ . Now,  $w$  is different from all  $v_j$  with  $0 \leq j < s-1$ , since otherwise  $(v_j, v_{j+1}, \dots, v_s, w = v_j)$  would be a cycle in  $G$ . But this means  $(v_0, v_1, \dots, v_s, w)$  is a path in  $G$  longer than  $P$ , contradicting maximality of  $P$ . So  $v_s$  must be a leaf. A similar argument shows that  $v_0$  is also a leaf. □

## 3.5 Functional Digraphs

We can obtain structural information about functions  $f : V \rightarrow V$  by viewing these functions as certain kinds of digraphs.

**3.38. Definition: Functional Digraphs.** A *functional digraph* on  $V$  is a simple digraph  $G$  with vertex set  $V$  such that  $\text{outdeg}_G(v) = 1$  for all  $v \in V$ .

**FIGURE 3.5**

A functional digraph.

A function  $f : V \rightarrow V$  can be thought of as a set  $E_f$  of ordered pairs such that for each  $x \in V$ , there exists exactly one  $y \in V$  with  $(x, y) \in E_f$ , namely  $y = f(x)$ . Then  $(V, E_f)$  is a functional digraph on  $V$ . Conversely, a functional digraph  $G = (V, E)$  determines a unique function  $g : V \rightarrow V$  by letting  $g(v)$  be the other endpoint of the unique edge in  $E$  departing from  $v$ . These comments establish a bijection between the set of functions on  $V$  and the set of functional digraphs on  $V$ .

**3.39. Example.** Figure 3.5 displays the functional digraph associated to the following function:

$$\begin{aligned} f(1) &= 15; & f(2) &= 16; & f(3) &= 8; & f(4) &= 17; & f(5) &= 5; \\ f(6) &= 5; & f(7) &= 4; & f(8) &= 3; & f(9) &= 6; & f(10) &= 4; \\ f(11) &= 10; & f(12) &= 4; & f(13) &= 10; & f(14) &= 1; & f(15) &= 12; \\ f(16) &= 1; & f(17) &= 15. \end{aligned}$$

We wish to understand the structure of functional digraphs. Consider the digraph  $G = (V, E)$  shown in Figure 3.5. Some of the vertices in this digraph are involved in cycles, which are drawn at the bottom of the figure. These cycles have length one or greater, and any two distinct cycles involve disjoint sets of vertices. The other vertices in the digraph all feed into these cycles at different points. We can form a set partition of the vertex set of the digraph by collecting together all vertices that feed into a particular vertex on a particular cycle. Each such collection can be viewed as a smaller digraph that has no cycles. We will show that these observations hold for all functional digraphs. To do this, we need a few more definitions.

**3.40. Definition: Cyclic Vertices.** Let  $G$  be a functional digraph on  $V$ . A vertex  $v \in V$  is called *cyclic* iff  $v$  belongs to some cycle of  $G$ ; otherwise,  $v$  is called *acyclic*.

**3.41. Example.** The cyclic elements for the functional digraph in Figure 3.5 are 3, 4, 5, 8, 12, 15, and 17.

Let  $f : V \rightarrow V$  be the function associated to the functional digraph  $G$ . Then  $v \in V$  is cyclic iff  $f^i(v) = v$  for some  $i \geq 1$ , where  $f^i$  denotes the composition of  $f$  with itself  $i$  times. This fact follows since the only possible cycle involving  $v$  in  $G$  must look like  $(v, f(v), f^2(v), f^3(v), \dots)$ .

**3.42. Definition: Rooted Trees.** A digraph  $G$  is called a *rooted tree with root  $v_0$*  iff  $G$  is a functional digraph and  $v_0$  is the unique cyclic vertex of  $G$ .

**3.43. Theorem: Structure of Functional Digraphs.** Let  $G$  be a functional digraph on  $V$  with associated function  $f : V \rightarrow V$ . Let  $C \subseteq V$  denote the set of cyclic vertices of  $G$ .  $C$  is nonempty, and each  $v \in C$  belongs to exactly one cycle of  $G$ . Also, there exists a unique indexed set partition  $\{S_v : v \in C\}$  of  $V$  such that the following hold for all  $v \in C$ : (i)  $v \in S_v$ ; (ii)  $x \in S_v$  and  $x \neq v$  implies  $f(x) \in S_v$ ; (iii) if  $g : S_v \rightarrow S_v$  is defined by  $g(x) = f(x)$  for  $x \neq v$ ,  $g(v) = v$ , then the functional digraph of  $g$  is a rooted tree with root  $v$ .

*Proof.* First, suppose  $v \in C$ . Since every vertex of  $G$  has exactly one outgoing edge, the only possible cycle involving  $v$  must be  $(v, f(v), f^2(v), \dots, f^i(v) = v)$ . So each cyclic vertex (if any) belongs to a unique cycle of  $G$ . This implies that distinct cycles of  $G$  involve disjoint sets of vertices and edges.

Next we define a surjection  $r : V \rightarrow C$ . The existence of  $r$  will show that  $C \neq \emptyset$ , since  $V \neq \emptyset$ . Fix  $u \in V$ . By repeatedly following outgoing arrows, we obtain for each  $k \geq 0$  a unique walk  $(u = u_0, u_1, u_2, \dots, u_k)$  in  $G$  of length  $k$ . Since  $V$  is finite, there must exist  $i < j$  with  $u_i = u_j$ . Take  $i$  minimal and then  $j$  minimal with this property; then  $(u_i, u_{i+1}, \dots, u_j)$  is a cycle in  $G$ . We define  $r(u) = u_i$ , which is the first element on this cycle reached from  $u$ . One may check that  $r(u) = u$  for all  $u \in C$ ; this implies that  $r$  is surjective. On the other hand, if  $u \notin C$ , the definition of  $r$  shows that  $r(u) = r(u_1) = r(f(u))$ .

How shall we construct a set partition with the stated properties? For each  $v \in C$ , consider the “fiber”  $S_v = r^{-1}(\{v\}) = \{w \in V : r(w) = v\}$ ; then  $\{S_v : v \in C\}$  is a set partition of  $V$  indexed by  $C$ . The remarks at the end of the last paragraph show that this set partition satisfies (i) and (ii). To check (iii) for some  $v \in C$ , first note that the map  $g$  defined in (iii) does map  $S_v$  into  $S_v$  by (i) and (ii). Suppose  $W = (w_0, w_1, \dots, w_k)$  is a cycle in the functional digraph for  $g$ . Since  $r(w_0) = v$ , we will eventually reach  $v$  by following outgoing arrows starting at  $w_0$ . On the other hand, following these arrows keeps us on the cycle  $W$ , so some  $w_i = v$ . Since  $g(v) = v$ , the only possibility is that  $W$  is the 1-cycle  $(v)$ . Thus (iii) holds for each  $v \in C$ .

To see that  $\{S_v : v \in C\}$  is unique, let  $\mathcal{P} = \{T_v : v \in C\}$  be another set partition with properties (i), (ii), and (iii). It is enough to show that  $S_v \subseteq T_v$  for each  $v \in C$ . Fix  $v \in C$  and  $z \in S_v$ . By (ii), every element in the sequence  $(z, f(z), f^2(z), \dots)$  belongs to the same set of  $\mathcal{P}$ , say  $T_w$ . Then  $v = r(z) = f^i(z) \in T_w$ , so (i) forces  $w = v$ . Thus  $z \in T_v$  as desired.  $\square$

We can informally summarize the previous result by saying that *every functional digraph uniquely decomposes into disjoint rooted trees feeding into one or more disjoint cycles*. There are two extreme cases of this decomposition that are especially interesting — the case where there are no trees (i.e.,  $C = V$ ), and the case where the whole digraph is a rooted tree (i.e.,  $|C| = 1$ ). We study these types of functional digraphs in the next two sections.

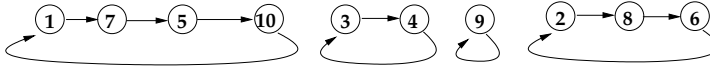
## 3.6 Cycle Structure of Permutations

The functional digraph of a bijection (permutation) has a particularly nice structure.

**3.44. Example.** Figure 3.6 displays the digraph associated to the following bijection:

$$\begin{aligned} h(1) &= 7; & h(2) &= 8; & h(3) &= 4; & h(4) &= 3; & h(5) &= 10; \\ h(6) &= 2; & h(7) &= 5; & h(8) &= 6; & h(9) &= 9; & h(10) &= 1. \end{aligned}$$

We see that the digraph for  $h$  contains only cycles; there are no trees feeding into these cycles. To see why this happens, compare this digraph to the digraph for the non-bijective

**FIGURE 3.6**

Digraph associated to a permutation.

function  $f$  in Figure 3.5. The digraph for  $f$  has a rooted tree feeding into the cyclic vertex 15. Accordingly,  $f$  is not injective since  $f(17) = 15 = f(1)$ . Similarly, if we move backwards through the trees in the digraph of  $f$ , we reach vertices with indegree zero (namely 2, 7, 9, 11, 13, and 14). The existence of such vertices shows that  $f$  is not surjective. Returning to the digraph of  $h$ , consider what happens if we reverse the direction of all the edges in the digraph. We obtain another functional digraph corresponding to the following function:

$$\begin{aligned} h'(1) &= 10; & h'(2) &= 6; & h'(3) &= 4; & h'(4) &= 3; & h'(5) &= 7; \\ h'(6) &= 8; & h'(7) &= 1; & h'(8) &= 2; & h'(9) &= 9; & h'(10) &= 5. \end{aligned}$$

One sees immediately that  $h'$  is the two-sided inverse for  $h$ .

The next theorem explains the observations in the last example.

**3.45. Theorem: Cycle Decomposition of Permutations.** Let  $f : V \rightarrow V$  be a function with functional digraph  $G$ . The map  $f$  is a bijection iff every  $v \in V$  is a cyclic vertex in  $V$ . In this situation,  $G$  is a disjoint union of cycles.

*Proof.* Suppose  $u \in V$  is a non-cyclic vertex. By 3.43,  $u$  belongs to a rooted tree  $S_v$  whose root  $v$  belongs to a cycle of  $G$ . Following edges outward from  $u$  will eventually lead to  $v$ ; let  $y$  be the vertex in  $S_v$  just before  $v$  on this path. Let  $z$  be the vertex just before  $v$  in the unique cycle involving  $v$ . We have  $y \neq z$ , but  $f(y) = v = f(z)$ . Thus,  $f$  is not injective.

Conversely, suppose all vertices in  $V$  are cyclic. Then the digraph  $G$  is a disjoint union of directed cycles. So every  $v \in V$  has indegree 1 as well as outdegree 1. Reversing the direction of every edge in  $G$  therefore produces another *functional* digraph  $G'$ . Let  $f' : V \rightarrow V$  be the function associated to this new digraph. For all  $a, b \in V$ , we have  $b = f(a)$  iff  $(a, b) \in G$  iff  $(b, a) \in G'$  iff  $a = f'(b)$ . It follows that  $f'$  is the two-sided inverse for  $f$ , so that  $f$  and  $f'$  are both bijections.  $\square$

Recall that  $S(n, k)$ , the Stirling number of the second kind, is the number of set partitions of an  $n$ -element set into  $k$  blocks. Let  $c(n, k)$  be the number of permutations of an  $n$ -element set whose functional digraph consists of  $k$  disjoint cycles. We will show that  $c(n, k) = s'(n, k)$ , the signless Stirling number of the first kind. Recall from 2.66 that the numbers  $s'(n, k)$  satisfy the recursion  $s'(n, k) = s'(n-1, k-1) + (n-1)s'(n-1, k)$  for  $0 < k < n$ , with initial conditions  $s'(n, 0) = \chi(n = 0)$  and  $s'(n, n) = 1$ .

**3.46. Theorem: Recursion for  $c(n, k)$ .** We have  $c(n, 0) = \chi(n = 0)$  and  $c(n, n) = 1$  for all  $n \geq 0$ . For  $0 < k < n$ , we have

$$c(n, k) = c(n-1, k-1) + (n-1)c(n-1, k).$$

Therefore,  $c(n, k) = s'(n, k)$  for all  $0 \leq k \leq n$ .

*Proof.* The identity map is the unique permutation of an  $n$ -element set with  $n$  cycles (which must each have length 1), so  $c(n, n) = 1$ . The only permutation with zero cycles is the empty function on the empty set, so  $c(n, 0) = \chi(n = 0)$ . Now suppose  $0 < k < n$ . Let  $A, B, C$  be

the sets of permutations counted by  $c(n, k)$ ,  $c(n-1, k-1)$ , and  $c(n-1, k)$ , respectively. Note that  $A$  is the disjoint union of the two sets

$$A_1 = \{f \in A : f(n) = n\} \text{ and } A_2 = \{f \in A : f(n) \neq n\}.$$

For each  $f \in A_1$ , we can restrict  $f$  to the domain  $\{1, 2, \dots, n-1\}$  to obtain a permutation of these  $n-1$  elements. Since  $f$  has  $k$  cycles, one of which involves  $n$  alone, the restriction of  $f$  must have  $k-1$  cycles. Since  $f \in A_1$  is uniquely determined by its restriction to  $\{1, 2, \dots, n-1\}$ , we have a bijection from  $A_1$  onto  $B$ .

On the other hand, let us build a typical element  $f \in A_2$  by making two choices. First, choose a permutation  $g \in C$  in  $c(n-1, k)$  ways. Second, choose an element  $i \in \{1, 2, \dots, n-1\}$  in  $n-1$  ways. Let  $j$  be the unique number such that  $g(j) = i$ . Modify the digraph for  $g$  by removing the arrow from  $j$  to  $i$  and replacing it by an arrow from  $j$  to  $n$  and an arrow from  $n$  to  $i$ . Informally, we are splicing  $n$  into the cycle just before  $i$ . Let  $f$  be the permutation associated to the new digraph. Evidently, the splicing process does not change the number of cycles of  $g$ , and  $f$  satisfies  $f(n) \neq n$ . Thus,  $f \in A_2$ , and every element of  $A_2$  arises uniquely by the choice process we have described. By the sum and product rules,

$$c(n, k) = |A| = |A_1| + |A_2| = c(n-1, k-1) + (n-1)c(n-1, k).$$

So  $c(n, k)$  and  $s'(n, k)$  satisfy the same recursion and initial conditions. A routine induction argument now shows that  $c(n, k) = s'(n, k)$  for all  $n$  and  $k$ .  $\square$

### 3.7 Counting Rooted Trees

Our goal in this section is to count rooted trees (see 3.42) with a fixed root vertex.

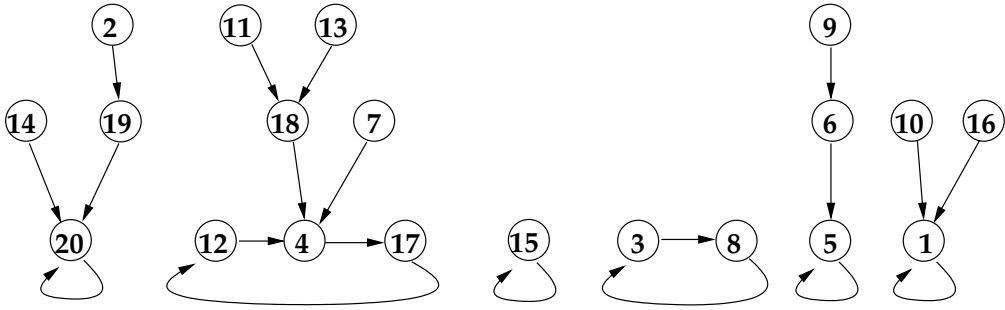
**3.47. Theorem: Enumeration of Rooted Trees.** For all  $n > 1$ , there are  $n^{n-2}$  rooted trees on the vertex set  $\{1, 2, \dots, n\}$  with root 1.

*Proof.* Let  $B$  be the set of rooted trees mentioned in the theorem. Let  $A$  be the set of all functions  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  such that  $f(1) = 1$  and  $f(n) = n$ . The product rule shows that  $|A| = n^{n-2}$ . It therefore suffices to define maps  $\phi : A \rightarrow B$  and  $\phi' : B \rightarrow A$  that are mutual inverses. To define  $\phi$ , fix  $f \in A$ . Let  $G_f = (\{1, 2, \dots, n\}, \{(i, f(i)) : 1 \leq i \leq n\})$  be the functional digraph associated with  $f$ . By 3.43, we can decompose the vertex set  $\{1, 2, \dots, n\}$  of this digraph into some disjoint cycles  $C_0, C_1, \dots, C_k$  and (possibly) some trees feeding into these cycles. For  $0 \leq i \leq k$ , let  $\ell_i$  be the largest vertex in cycle  $C_i$ , and write  $C_i = (r_i, \dots, \ell_i)$ . We can choose the indexing of the cycles so that the numbers  $\ell_i$  satisfy  $\ell_0 > \ell_1 > \ell_2 > \dots > \ell_k$ . Since  $f(1) = 1$  and  $f(n) = n$ , 1 and  $n$  belong to cycles of length 1, so that  $\ell_0 = r_0 = n$ ,  $C_0 = (n)$ ,  $\ell_k = r_k = 1$ ,  $C_k = (1)$ , and  $k > 0$ . To obtain  $\phi(f)$ , modify the digraph  $G_f$  by removing all edges of the form  $(\ell_i, r_i)$  and adding new edges  $(\ell_i, r_{i+1})$ , for  $0 \leq i < k$ . One may check that  $\phi(f)$  is always a rooted tree with root 1.

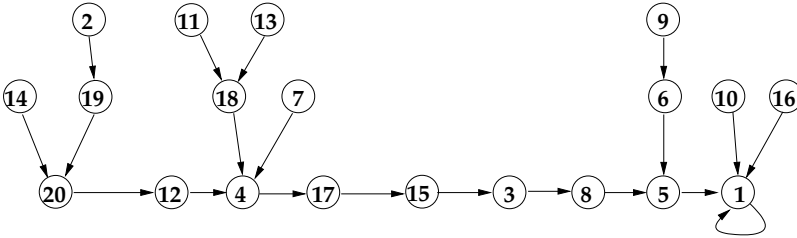
**3.48. Example.** Suppose  $n = 20$  and  $f$  is the function defined as follows:

$$\begin{array}{lllll} f(1) = 1; & f(2) = 19; & f(3) = 8; & f(4) = 17; & f(5) = 5; \\ f(6) = 5; & f(7) = 4; & f(8) = 3; & f(9) = 6; & f(10) = 1; \\ f(11) = 18; & f(12) = 4; & f(13) = 18; & f(14) = 20; & f(15) = 15; \\ f(16) = 1; & f(17) = 12; & f(18) = 4; & f(19) = 20; & f(20) = 20. \end{array}$$

We draw the digraph of  $f$  in such a way that all vertices involved in cycles occur in a horizontal row at the bottom of the figure, and the largest element in each cycle is the rightmost

**FIGURE 3.7**

A functional digraph with cycles arranged in canonical order.

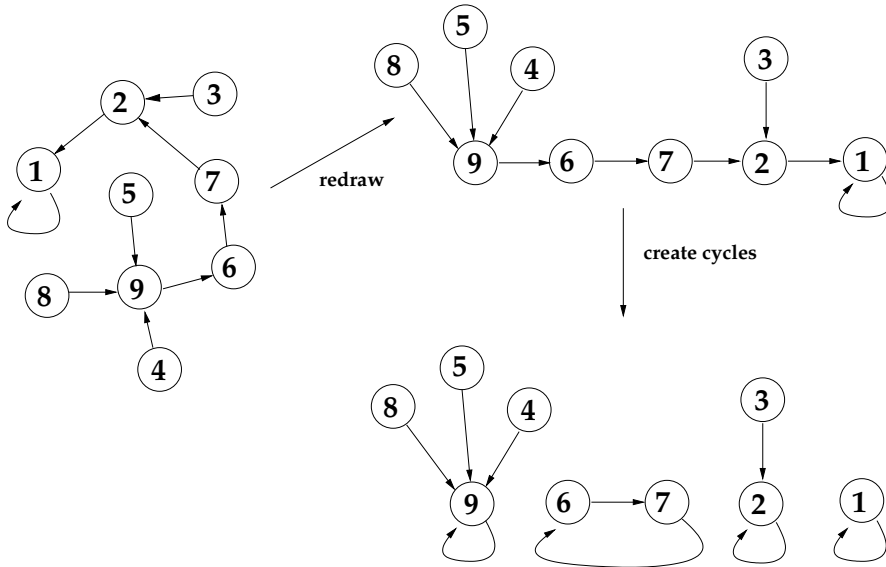
**FIGURE 3.8**

Conversion of the digraph to a rooted tree.

element of its cycle. We arrange these cycles so that these largest elements decrease from left to right; in particular, vertex  $n$  is always at the far left, and vertex 1 at the far right. See Figure 3.7. To compute  $\phi(f)$ , we cut the “back-edges” leading left from  $\ell_i$  to  $r_i$  (which are loops if  $\ell_i = r_i$ ) and add new edges leading right from  $\ell_i$  to  $r_{i+1}$ . See Figure 3.8.

Continuing the proof, let us see why  $\phi$  is invertible. Let  $T$  be a rooted tree on  $\{1, 2, \dots, n\}$  with root 1. Following outgoing edges from  $n$  must eventually lead to the unique cyclic vertex 1. Let  $P = (v_0, v_1, \dots, v_s)$  be the vertices encountered on the way from  $v_0 = n$  to  $v_s = 1$ . We recursively recover the numbers  $\ell_0, \ell_1, \dots, \ell_k$  as follows. Let  $\ell_0 = n$ . Define  $\ell_1$  to be the largest number in  $P$  following  $\ell_0$ . In general, after  $\ell_{i-1}$  has been found, define  $\ell_i$  to be the largest number in  $P$  following  $\ell_{i-1}$ . After finitely many steps, we will get  $\ell_k = 1$  for some  $k$ . Next, let  $r_0 = n$ , and for  $i > 0$ , let  $r_i$  be the vertex immediately following  $\ell_{i-1}$  on the path  $P$ . Modify  $T$  by deleting the edges  $(\ell_i, r_{i+1})$  and adding edges of the form  $(\ell_i, r_i)$ , for  $0 \leq i < k$ . One can verify that every vertex in the resulting digraph  $G'$  has outdegree exactly 1, and there are loop edges in  $G'$  at vertex 1 and vertex  $n$ . Thus,  $G'$  is a functional digraph that determines a function  $f = \phi'(T) \in A$ . It follows from the definition of  $\phi'$  that  $\phi'$  is the two-sided inverse of  $\phi$ .  $\square$

**3.49. Example.** Suppose  $n = 9$  and  $T$  is the rooted tree shown on the left in Figure 3.9. We first redraw the picture of  $T$  so that the vertices on the path  $P$  from  $n$  to 1 occur in a horizontal row at the bottom of the picture, with  $n$  on the left and 1 on the right. We recover  $\ell_i$  and  $r_i$  by the procedure above, and then delete the appropriate edges of  $T$  and add appropriate back-edges to create cycles. The resulting functional digraph appears on

**FIGURE 3.9**

Conversion of a rooted tree to a functional digraph.

the bottom right in Figure 3.9. So  $\phi'(T)$  is the function  $g$  defined as follows:

$$\begin{aligned} g(1) &= 1; & g(2) &= 2; & g(3) &= 2; & g(4) &= 9; & g(5) &= 9; \\ g(6) &= 7; & g(7) &= 6; & g(8) &= 9; & g(9) &= 9. \end{aligned}$$

### 3.8 Connectedness and Components

In many applications of graphs, it is important to know whether every vertex is reachable from every other vertex.

**3.50. Definition: Connectedness.** Let  $G = (V, E, \epsilon)$  be a graph or digraph.  $G$  is *connected* iff for all  $u, v \in V$ , there is a walk in  $G$  from  $u$  to  $v$ .

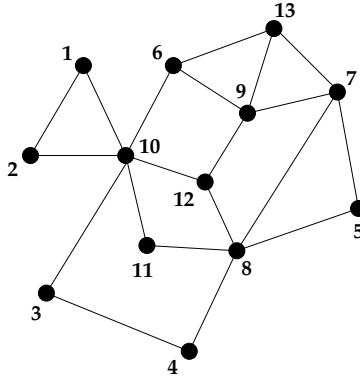
**3.51. Example.** The graph  $G_1$  in Figure 3.1 is connected, but the simple graph  $G_2$  and the digraph  $G_3$  in that figure are not connected.

Connectedness can also be described using paths instead of walks.

**3.52. Theorem: Walks vs. Paths.** Let  $G = (V, E, \epsilon)$  be a graph or digraph, and let  $u, v \in V$ . There is a walk in  $G$  from  $u$  to  $v$  iff there is a path in  $G$  from  $u$  to  $v$ .

*Proof.* Let  $W = (v_0, e_1, v_1, \dots, e_s, v_s)$  be a walk in  $G$  from  $u$  to  $v$ . We describe an algorithm to convert the walk  $W$  into a path from  $u$  to  $v$ . If all the vertices  $v_i$  are distinct, then the edges  $e_i$  must also be distinct, so  $W$  is already a path. Otherwise, choose  $i$  minimal such that  $v_i$  appears more than once in  $W$ , and then choose  $j$  maximal such that  $v_i = v_j$ . Then  $W_1 = (v_0, e_1, v_1, \dots, e_i, v_i, e_{j+1}, v_{j+1}, \dots, e_s, v_s)$  is a walk from  $u$  to  $v$  of shorter length than



**FIGURE 3.10**

Converting a walk to a path.

$W$ . If  $W_1$  is a path, we are done. Otherwise, we repeat the argument to obtain a walk  $W_2$  from  $u$  to  $v$  that is shorter than  $W_1$ . Since the lengths keep decreasing, this process must eventually terminate with a path  $W_k$  from  $u$  to  $v$ . ( $W_k$  has length zero if  $u = v$ .) The converse is immediate, since every path in  $G$  from  $u$  to  $v$  is a walk in  $G$  from  $u$  to  $v$ .  $\square$

**3.53. Example.** In the simple graph shown in Figure 3.10, consider the walk

$$W = (11, 10, 1, 2, 10, 3, 4, 8, 11, 8, 12, 10, 6, 9, 7, 13, 9, 12, 8, 5).$$

First, the repetition  $v_0 = 11 = v_8$  leads to the walk

$$W_1 = (11, 8, 12, 10, 6, 9, 7, 13, 9, 12, 8, 5).$$

Eliminating the multiple visits to vertex 8 leads to the walk

$$W_2 = (11, 8, 5).$$

$W_2$  is a path from 11 to 5.

**3.54. Corollary: Connectedness and Paths.** A graph or digraph  $G = (V, E, \epsilon)$  is connected iff for all  $u, v \in V$ , there is *at least one path* in  $G$  from  $u$  to  $v$ .

By looking at pictures of graphs, it becomes visually evident that any graph decomposes into a disjoint union of connected pieces, with no edge joining vertices in two separate pieces. These pieces are called the (connected) components of the graph. The situation for digraphs is more complicated, since there may exist directed edges between different components. To give a formal development of these ideas, we introduce the following equivalence relation.

**3.55. Definition: Interconnection Relation.** Let  $G = (V, E, \epsilon)$  be a graph or digraph. Define a binary relation  $\leftrightarrow_G$  on the vertex set  $V$  by setting  $u \leftrightarrow_G v$  iff there exist walks in  $G$  from  $u$  to  $v$  and from  $v$  to  $u$ .

In the case of *graphs*, note that  $u \leftrightarrow_G w$  iff there is a walk in  $G$  from  $u$  to  $w$ , since the reversal of such a walk is a walk in  $G$  from  $w$  to  $u$ . Now, for a graph or digraph  $G$ , let us verify that  $\leftrightarrow_G$  is indeed an equivalence relation on  $V$ . First, for all  $u \in V$ ,  $(u)$  is a walk of length 0 from  $u$  to  $u$ , so  $u \leftrightarrow_G u$  and  $\leftrightarrow_G$  is reflexive on  $V$ . Second, the symmetry of  $\leftrightarrow_G$  is automatic from the way we defined  $\leftrightarrow_G$ :  $u \leftrightarrow_G v$  implies  $v \leftrightarrow_G u$  for all  $u, v \in V$ . Finally,

to check transitivity, suppose  $u, v, w \in V$  satisfy  $u \leftrightarrow_G v$  and  $v \leftrightarrow_G w$ . Let  $W_1, W_2, W_3$ , and  $W_4$  be walks in  $G$  from  $u$  to  $v$ , from  $v$  to  $w$ , from  $v$  to  $u$ , and from  $w$  to  $v$ , respectively. Then the concatenation of  $W_1$  followed by  $W_2$  is a walk in  $G$  from  $u$  to  $w$ , whereas the concatenation of  $W_4$  followed by  $W_3$  is a walk in  $G$  from  $w$  to  $u$ . Hence  $u \leftrightarrow_G w$ , as desired.

**3.56. Definition: Components.** Let  $G = (V, E, \epsilon)$  be a graph or digraph. The *components* of  $G$  are the equivalence classes of the interconnection equivalence relation  $\leftrightarrow_G$ . Components are also called *connected components* or (in the case of digraphs) *strong components*.

Since  $\leftrightarrow_G$  is an equivalence relation on  $V$ , the components of  $G$  form a set partition of the vertex set  $V$ . Given a component  $C$  of  $G$ , consider the graph or digraph  $(C, E', \epsilon')$  obtained by retaining those edges in  $E$  with both endpoints in  $C$  and restricting  $\epsilon$  to this set of edges. One may check that this graph or digraph is connected.

**3.57. Example.** The components of the graph  $G_2$  in Figure 3.1 are  $\{0\}$  and  $\{1, 2, 3, 4, 5, 6\}$ . The components of the digraph  $G_3$  in that figure are  $\{1, 2, 4, 5\}$ ,  $\{3\}$ ,  $\{6\}$ , and  $\{7\}$ .

The next theorems describe how the addition or deletion of an edge affects the components of a graph.

**3.58. Theorem: Edge Deletion and Components.** Let  $G = (V, E, \epsilon)$  be a graph with components  $\{C_i : i \in I\}$ . Let  $e \in E$  be an edge with endpoints  $v, w \in C_j$ . Let  $G' = (V, E', \epsilon')$  where  $E' = E \sim \{e\}$  and  $\epsilon'$  is the restriction of  $\epsilon$  to  $E'$ .

- (a) If  $e$  appears in some cycle of  $G$ , then  $G$  and  $G'$  have the same components.
- (b) If  $e$  appears in no cycle of  $G$ , then  $G'$  has one more component than  $G$ . More precisely, the components of  $G'$  are the  $C_k$  with  $k \neq j$ , together with two disjoint sets  $A$  and  $B$  such that  $A \cup B = C_j$ ,  $v \in A$ , and  $w \in B$ .

*Proof.* For (a), let  $(v_0, e_1, v_1, e_2, \dots, v_s)$  be a cycle of  $G$  containing  $e$ . Cyclically shifting and reversing the cycle if needed, we can assume  $v_0 = v = v_s$ ,  $e_1 = e$ , and  $v_1 = w$ . Statement (a) will follow if we can show that the interconnection relations  $\leftrightarrow_G$  and  $\leftrightarrow_{G'}$  coincide. First, for all  $y, z \in V$ ,  $y \leftrightarrow_{G'} z$  implies  $y \leftrightarrow_G z$  since every walk in the smaller graph  $G'$  is also a walk in  $G$ . On the other hand, does  $y \leftrightarrow_G z$  imply  $y \leftrightarrow_{G'} z$ ? We know there is a walk  $W$  from  $y$  to  $z$  in  $G$ . If  $W$  does not use the edge  $e$ ,  $W$  is a walk from  $y$  to  $z$  in  $G'$ . Otherwise, we can modify  $W$  as follows. Every time  $W$  goes from  $v = v_s = v_0$  to  $w = v_1$  via  $e$ , replace this part of the walk by the sequence  $(v_s, e_s, \dots, e_2, v_1)$  obtained by taking a detour around the cycle. Make a similar modification each time  $W$  goes from  $w$  to  $v$  via  $e$ . This produces a walk in  $G'$  from  $y$  to  $z$ .

For (b), let us compute the equivalence classes of  $\leftrightarrow_{G'}$ . First, fix  $z \in C_k$  where  $k \neq j$ . The set  $C_k$  consists of all vertices in  $V$  reachable from  $z$  by walks in  $G$ . One readily checks that none of these walks can use the edge  $e$ , so  $C_k$  is also the set of all vertices in  $V$  reachable from  $z$  by walks in  $G'$ . So  $C_k$  is the equivalence class of  $z$  relative to both  $\leftrightarrow_G$  and  $\leftrightarrow_{G'}$ .

Next, let  $A$  and  $B$  be the equivalence classes of  $v$  and  $w$  (respectively) relative to  $\leftrightarrow_{G'}$ . By definition,  $A$  and  $B$  are two of the components of  $G'$  (possibly the same component). We now show that  $A$  and  $B$  are disjoint and that their union is  $C_j$ . If the equivalence classes  $A$  and  $B$  are not disjoint, then they must be equal. By 3.52, there must be a path  $(v_0, e_1, v_1, \dots, e_s, v_s)$  in  $G'$  from  $v$  to  $w$ . Appending  $e, v_0$  to this path would produce a cycle in  $G$  involving the edge  $e$ , which is a contradiction. Thus  $A$  and  $B$  are disjoint. Let us show that  $A \cup B \subseteq C_j$ . If  $z \in A$ , then there is a walk in  $G'$  (and hence in  $G$ ) from  $v$  to  $z$ . Since  $C_j$  is the equivalence class of  $v$  relative to  $\leftrightarrow_G$ , it follows that  $z \in C_j$ . Similarly,  $z \in B$  implies  $z \in C_j$  since  $C_j$  is also the equivalence class of  $w$  relative to  $\leftrightarrow_G$ . Next, we check that  $C_j \subseteq A \cup B$ . Let  $z \in C_j$ , and let  $W = (w_0, e_1, w_1, \dots, w_t)$  be a walk in  $G$  from  $v$  to  $z$ . If  $W$  does not use the edge  $e$ , then  $z \in A$ . If  $W$  does use  $e$ , then the portion of  $W$  following the

last appearance of the edge  $e$  is a walk from either  $v$  or  $w$  to  $z$  in  $G'$ ; thus  $z \in A \cup B$ . Since the union of  $A$ ,  $B$ , and the  $C_k$  with  $k \neq j$  is all of  $V$ , we have found all the components of  $G'$ .  $\square$

The previous result suggests the following terminology.

**3.59. Definition: Cut-Edges.** An edge  $e$  in a graph  $G$  is a *cut-edge* iff  $e$  does not appear in any cycle of  $G$ .

**3.60. Theorem: Edge Addition and Components.** Let  $G = (V, E, \epsilon)$  be a graph with components  $\{C_i : i \in I\}$ . Let  $G^+ = (V, E^+, \epsilon^+)$  be the graph obtained from  $G$  by adding a new edge  $e$  with endpoints  $v \in C_j$  and  $w \in C_k$ .

(a) If  $v$  and  $w$  are in the same component  $C_j$  of  $G$ , then  $e$  is involved in a cycle of  $G^+$ , and  $G$  and  $G^+$  have the same components.

(b) If  $v$  and  $w$  are in different components of  $G$ , then  $e$  is a cut-edge of  $G^+$ , and the components of  $G^+$  are  $C_j \cup C_k$  and the  $C_i$  with  $i \neq j, k$ .

This theorem follows readily from 3.58, so we leave the proof to the reader.

### 3.9 Forests

**3.61. Definition: Forests.** A *forest* is a graph with no cycles. Such a graph is also called *acyclic*.

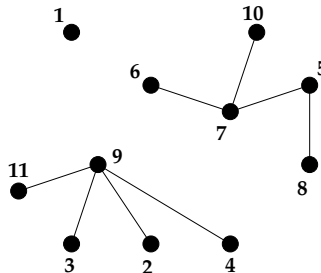
A forest cannot have any loops or multiple edges between the same two vertices. So we can assume, with no loss of generality, that forests are *simple* graphs.

**3.62. Example.** Figure 3.11 displays a forest.

Recall from 3.54 that a graph  $G$  is connected iff there exists *at least one* path between any two vertices of  $G$ . The next result gives an analogous characterization of forests.

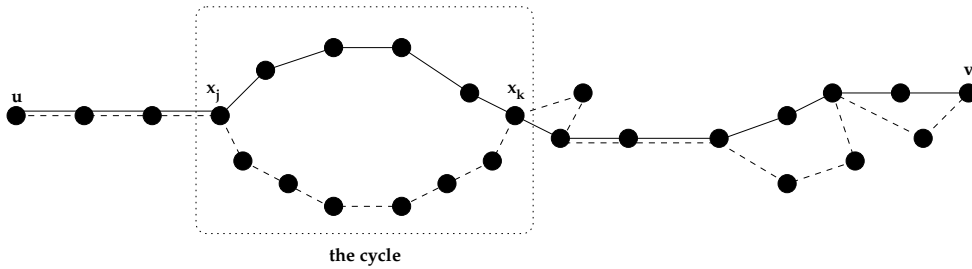
**3.63. Theorem: Forests and Paths.** A graph  $G$  is acyclic iff  $G$  has no loops and for all  $u, v$  in  $V(G)$ , there is *at most one* path from  $u$  to  $v$  in  $G$ .

*Proof.* We prove the contrapositive in both directions. First suppose that  $G$  has a cycle  $C = (v_0, e_1, v_1, \dots, e_s, v_s)$ . If  $s = 1$ ,  $G$  has a loop. If  $s > 1$ , then  $(v_1, e_2, \dots, e_s, v_s)$  and



**FIGURE 3.11**

A forest.

**FIGURE 3.12**

Extracting a cycle from two paths.

$(v_1, e_1, v_0)$  are two distinct paths in  $G$  from  $v_1$  to  $v_0$ . For the converse, we can assume  $G$  is simple. Suppose  $u$  and  $v$  are vertices of  $G$  and  $P = (x_0, x_1, \dots, x_s)$ ,  $Q = (y_0, y_1, \dots, y_t)$  are two distinct paths in  $G$  from  $u$  to  $v$ , where all  $x_i$ 's and  $y_j$ 's are in  $V(G)$ . We will use these paths to construct a cycle in  $G$ . Note that the concatenation of  $P$  and the reversal of  $Q$  is a walk in  $G$  that starts and ends at  $u$ . But this walk need not be a cycle, since it may involve repeated edges or vertices.

Since  $P$  and  $Q$  are distinct, there is an index  $j$  such that  $x_i = y_i$  for  $0 \leq i \leq j$ , but  $x_{j+1} \neq y_{j+1}$ . Since  $P$  and  $Q$  both end at  $v$ , there must exist a least index  $k \geq j + 1$  such that  $x_k$  is a vertex in  $Q$ , say  $x_k = y_r$ . It follows from the choice of  $j$  and  $k$  that either  $k = j + 1$  and  $r > j + 1$ , or  $k > j + 1$  and  $r \geq j + 1$ . In any case,

$$C = (x_j, x_{j+1}, \dots, x_k = y_r, y_{r-1}, \dots, y_{j+1}, y_j)$$

is a cycle in  $G$ . Figure 3.12 illustrates this argument. □

The following result gives a formula for the number of components in a forest.

**3.64. Theorem: Components of a Forest.** Let  $G$  be a forest with  $n$  vertices and  $k$  edges. The number of connected components of  $G$  is  $n - k$ .

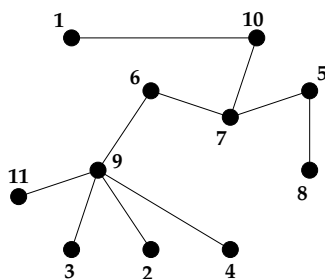
*Proof.* We use induction on  $k$ . The result holds for  $k = 0$ , since  $G$  consists of  $n$  isolated vertices in this case. Assume that  $k > 0$  and the result is already known for forests with  $n$  vertices and  $k - 1$  edges. Given a forest  $G$  with  $n$  vertices and  $k$  edges, remove one edge  $e$  from  $G$  to get a new graph  $H$ . The graph  $H$  is acyclic and has  $n$  vertices and  $k - 1$  edges. By induction,  $H$  has  $n - (k - 1) = n - k + 1$  components. On the other hand,  $e$  must be a cut-edge since  $G$  has no cycles. It follows from 3.58 that  $H$  has one more component than  $G$ . Thus,  $G$  has  $n - k$  components, as desired. □

### 3.10 Trees

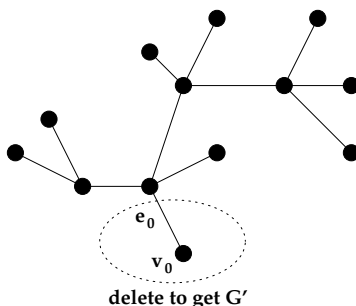
**3.65. Definition: Trees.** A *tree* is a connected graph with no cycles.

**3.66. Example.** Figure 3.13 displays a tree.

Every component of a forest is a tree, so every forest is a disjoint union of trees. The next result is an immediate consequence of 3.37.

**FIGURE 3.13**

A tree.

**FIGURE 3.14**

Pruning a leaf from a tree gives another tree.

**3.67. Theorem: Trees Have Leaves.** If  $T$  is a tree with more than one vertex, then  $T$  has at least two leaves.

**3.68. Definition: Pruning.** Suppose  $G = (V, E)$  is a simple graph,  $v_0$  is a leaf in  $G$ , and  $e_0$  is the unique edge incident to the vertex  $v_0$ . The graph obtained by pruning  $v_0$  from  $G$  is the graph  $(V \sim \{v_0\}, E \sim \{e_0\})$ .

**3.69. Pruning Lemma.** If  $T$  is an  $n$ -vertex tree,  $v_0$  is a leaf of  $T$ , and  $T'$  is obtained from  $T$  by pruning  $v_0$ , then  $T'$  is a tree with  $n - 1$  vertices.

*Proof.* First,  $T$  has no cycles, and the deletion of  $v_0$  and the associated edge  $e_0$  will not create any cycles. So  $T'$  is acyclic. Second, let  $u, w \in V(T')$ . There is a path from  $u$  to  $w$  in  $T$ . Since  $u \neq v_0 \neq w$ , this path will not use the edge  $e_0$  or the vertex  $v_0$ . Thus there is a path from  $u$  to  $w$  in  $T'$ , so  $T'$  is connected.  $\square$

Figure 3.14 illustrates the pruning lemma.

To illustrate an application of pruning, we now prove a fundamental relationship between the number of vertices and edges in a tree (this also follows from 3.64).

**3.70. Theorem: Number of Edges in a Tree.** If  $G$  is a tree with  $n > 0$  vertices, then  $G$  has  $n - 1$  edges.

*Proof.* We argue by induction on  $n$ . If  $n = 1$ , then  $G$  must have  $0 = n - 1$  edges. Assume  $n > 1$  and that the result holds for trees with  $n - 1$  vertices. Let  $T$  be a tree with  $n$  vertices. We know that  $T$  has at least one leaf; let  $v_0$  be one such leaf. Let  $T'$  be the graph obtained

by pruning  $v_0$  from  $T$ . By the pruning lemma,  $T'$  is a tree with  $n - 1$  vertices. By induction,  $T'$  has  $n - 2$  edges. Hence,  $T$  has  $n - 1$  edges.  $\square$

**3.71. Theorem: Characterizations of Trees.** Let  $G$  be a graph with  $n$  vertices. The following conditions are logically equivalent:

- (a)  $G$  is a tree (i.e.,  $G$  is connected and acyclic).
  - (b)  $G$  is connected and has  $\leq n - 1$  edges.
  - (c)  $G$  is acyclic and has  $\geq n - 1$  edges.
  - (d)  $G$  has no loop edges, and for all  $u, v \in V(G)$ , there is a unique path in  $G$  from  $u$  to  $v$ .
- Moreover, when these conditions hold,  $G$  has  $n - 1$  edges.

*Proof.* First, (a) implies (b) and (a) implies (c) by 3.70. Second, (a) is equivalent to (d) by virtue of 3.54 and 3.63. Third, let us prove (b) implies (a). Assume  $G$  is connected with  $k \leq n - 1$  edges. If  $G$  has a cycle, delete one edge on some cycle of  $G$ . The resulting graph is still connected (by 3.58) and has  $k - 1$  edges. Continue to delete edges in this way, one at a time, until there are no cycles. If we deleted  $i$  edges total, the resulting graph is a tree with  $k - i \leq n - 1 - i$  edges and  $n$  vertices. By 3.70, we must have  $i = 0$  and  $k = n - 1$ . So no edges were deleted, and  $G$  itself is in fact a tree.

Fourth, let us prove (c) implies (a). Assume  $G$  is acyclic with  $k \geq n - 1$  edges. If  $G$  is not connected, add an edge joining two distinct components of  $G$ . The resulting graph is still acyclic (by 3.60) and has  $k + 1$  edges. Continue to add edges in this way, one at a time, until the graph becomes connected. If we added  $i$  edges total, the resulting graph is a tree with  $k + i \geq n - 1 + i$  edges and  $n$  vertices. By 3.70, we must have  $i = 0$  and  $k = n - 1$ . So no edges were added, and  $G$  itself is in fact a tree.  $\square$

### 3.11 Counting Trees

The next theorem, usually attributed to Cayley, counts  $n$ -vertex trees.

**3.72. Theorem: Enumeration of Trees.** For all  $n \geq 1$ , there are  $n^{n-2}$  trees with vertex set  $\{1, 2, \dots, n\}$ .

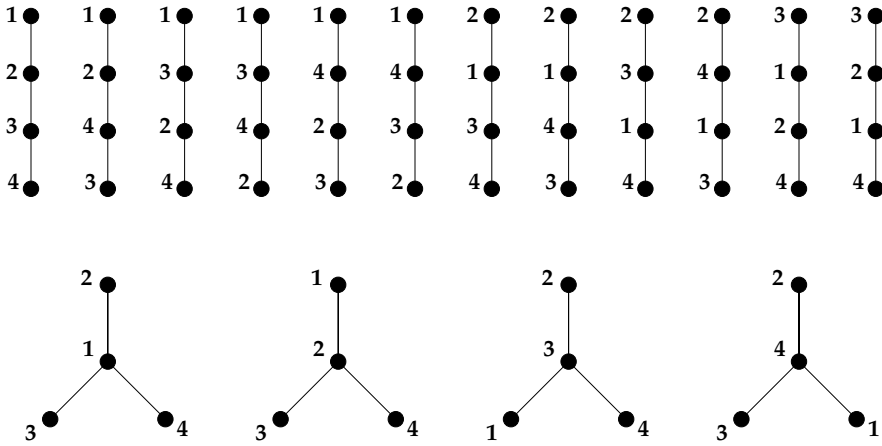
**3.73. Example.** Figure 3.15 displays all  $4^{4-2} = 16$  trees with vertex set  $\{1, 2, 3, 4\}$ .

Theorem 3.72 is an immediate consequence of 3.47 and the following bijection.

**3.74. Theorem: Trees vs. Rooted Trees.** Let  $V$  be a finite set and  $v_0 \in V$ . There is a bijection from the set  $A$  of trees with vertex set  $V$  to the set  $B$  of rooted trees with vertex set  $V$  and root  $v_0$ .

*Proof.* We define maps  $f : A \rightarrow B$  and  $g : B \rightarrow A$  that are two-sided inverses. First, given  $T = (V, E) \in A$ , construct  $f(T) = (V, E')$  as follows. For each  $v \in V$  with  $v \neq v_0$ , there exists a unique path from  $v$  to  $v_0$  in  $T$ . Letting  $e = \{v, w\}$  be the first edge on this path, we add the directed edge  $(v, w)$  to  $E'$ . Also, we add the loop edge  $(v_0, v_0)$  to  $E'$ . Since  $T$  has no cycles, the only possible cycle in the resulting functional digraph  $f(T)$  is the 1-cycle  $(v_0)$ . It follows that  $f(T)$  is a rooted tree on  $V$  with root  $v_0$  (see 3.42).

Next, given a rooted tree  $S \in B$ , define  $g(S)$  by deleting the unique loop edge  $(v_0, v_0)$  and replacing every directed edge  $(v, w)$  by an undirected edge  $\{v, w\}$ . The resulting graph  $g(S)$  has  $n$  vertices and  $n - 1$  edges. To see that  $g(S)$  is connected, fix  $y, z \in V$ . Following outgoing edges from  $y$  (resp.  $z$ ) in  $S$  produces a directed path from  $y$  (resp.  $z$ ) to  $v_0$  in  $S$ .

**FIGURE 3.15**

The 16 trees on four vertices.

In the undirected graph  $g(S)$ , we can concatenate the path from  $y$  to  $v_0$  with the reverse of the path from  $z$  to  $v_0$  to get a walk from  $y$  to  $z$ . It follows that  $g(S)$  is a tree.

It is routine to check that  $g \circ f = \text{id}_A$ , since  $f$  assigns a certain orientation to each edge of the original tree, and this orientation is then forgotten by  $g$ . It is somewhat less routine to verify that  $f \circ g = \text{id}_B$ ; we leave this to the reader. (One must check that the edge orientations in  $f(g(S))$  agree with the edge orientations in  $S$ , for each  $S \in B$ .)  $\square$

A different bijective proof of Cayley's theorem, which employs parking functions, is presented in §12.5. We next prove a refinement of Cayley's theorem that counts the number of trees such that each vertex has a specified degree. We give an algebraic proof first, and then convert this to a bijective proof in the next section.

**3.75. Theorem: Counting Trees with Specified Degrees.** Suppose  $n \geq 2$  and  $d_1, \dots, d_n \geq 0$  are fixed integers. If  $d_1 + \dots + d_n = 2n - 2$ , then there are

$$\binom{n-2}{d_1-1, d_2-1, \dots, d_n-1} = \frac{(n-2)!}{\prod_{j=1}^n (d_j-1)!}$$

trees with vertex set  $\{v_1, v_2, \dots, v_n\}$  such that  $\deg(v_j) = d_j$  for all  $j$ . If  $d_1 + \dots + d_n \neq 2n - 2$ , then there are no such trees.

*Proof.* The last statement holds because any tree  $T$  on  $n$  vertices has  $n - 1$  edges, and thus  $\sum_{i=1}^n \deg(v_i) = 2(n - 1)$ . Assume henceforth that  $d_1 + \dots + d_n = 2n - 2$ . We prove the result by induction on  $n$ . First consider the case  $n = 2$ . If  $d_1 = d_2 = 1$ , there is exactly one valid tree, and  $\binom{n-2}{d_1-1, d_2-1} = 1$ . For any other choice of  $d_1, d_2$  adding to 2, there are no valid trees, and  $\binom{n-2}{d_1-1, d_2-1} = 0$ .

Now assume  $n > 2$  and that the theorem is known to hold for trees with  $n - 1$  vertices. Let  $A$  be the set of trees  $T$  with  $V(T) = \{v_1, \dots, v_n\}$  and  $\deg(v_j) = d_j$  for all  $j$ . If  $d_j = 0$  for some  $j$ , then  $A$  is empty and the formula in the theorem is zero by convention. Now suppose  $d_j > 0$  for all  $j$ . We must have  $d_i = 1$  for some  $i$ , for otherwise  $d_1 + \dots + d_n \geq 2n > 2n - 2$ . Fix an  $i$  with  $d_i = 1$ . Note that  $v_i$  is a leaf in  $T$  for every  $T \in A$ . Now define

$$A_k = \{T \in A : \{v_i, v_k\} \in E(T)\} \quad (1 \leq k \leq n, k \neq i).$$

$A_k$  is the set of trees in  $A$  in which the leaf  $v_i$  is attached to the vertex  $v_k$ .  $A$  is the disjoint union of the sets  $A_k$ .

Fix  $k \neq i$ . Pruning the leaf  $v_i$  gives a bijection between  $A_k$  and the set  $B_k$  of all trees with vertex set  $\{v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n\}$  such that  $\deg(v_j) = d_j$  for  $j \neq i, k$  and  $\deg(v_k) = d_k - 1$ . By induction hypothesis,

$$|B_k| = \frac{(n-3)!}{(d_k-2)! \prod_{\substack{1 \leq j \leq n \\ j \neq i, k}} (d_j-1)!}.$$

Therefore,

$$\begin{aligned} |A| &= \sum_{k \neq i} |A_k| = \sum_{k \neq i} |B_k| = \sum_{k \neq i} \frac{(n-3)!}{(d_k-2)! \prod_{j \neq k, i} (d_j-1)!} \\ &= \sum_{k \neq i} \frac{(n-3)!(d_k-1)}{\prod_{j \neq i} (d_j-1)!} = \sum_{k \neq i} \frac{(n-3)!(d_k-1)}{\prod_{j=1}^n (d_j-1)!} \\ &\quad (\text{since } (d_i-1)! = 0! = 1) \\ &= \frac{(n-3)!}{\prod_{j=1}^n (d_j-1)!} \sum_{k \neq i} (d_k-1). \end{aligned}$$

Now, since  $d_i = 1$ ,  $\sum_{k \neq i} (d_k-1) = \sum_{k=1}^n (d_k-1) = (2n-2) - n = n-2$ . Inserting this into the previous formula, we see that

$$|A| = \frac{(n-2)!}{\prod_{j=1}^n (d_j-1)!},$$

which completes the induction proof.  $\square$

**3.76. Corollary: Second Proof of 3.72.** Let us sum the previous formula over all possible degree sequences  $(d_1, \dots, d_n)$ . Making the change of variables  $c_i = d_i - 1$  and invoking the multinomial theorem 2.12, we see that the total number of trees on this vertex set is

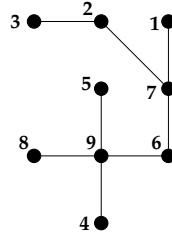
$$\begin{aligned} \sum_{\substack{d_1 + \dots + d_n = 2n-2 \\ d_i \geq 0}} \binom{n-2}{d_1-1, d_2-1, \dots, d_n-1} &= \sum_{\substack{c_1 + \dots + c_n = n-2 \\ c_i \geq 0}} \binom{n-2}{c_1, c_2, \dots, c_n} 1^{c_1} 1^{c_2} \dots 1^{c_n} \\ &= (1+1+\dots+1)^{n-2} = n^{n-2}. \end{aligned}$$

### 3.12 Pruning Maps

We now develop a bijective proof of 3.75. Suppose  $n \geq 2$  and  $d_1, \dots, d_n$  are positive integers that sum to  $2n-2$ . Let  $V = \{v_1, \dots, v_n\}$  be a vertex set consisting of  $n$  positive integers. Let  $A$  be the set of trees  $T$  with vertex set  $V$  such that  $\deg(v_i) = d_i$  for all  $i$ . Let  $B$  be the set of words  $\mathcal{R}(v_1^{d_1-1} v_2^{d_2-1} \dots v_n^{d_n-1})$  as in 1.44. Each word  $w \in B$  has length  $n-2$  and consists of  $d_j-1$  copies of  $v_j$  for all  $j$ . To prove 3.75, it suffices to define a bijection  $f: A \rightarrow B$ .

Given a tree  $T \in A$ , we compute  $f(T)$  by repeatedly pruning off the largest leaf of  $T$ , recording for each leaf the vertex adjacent to it in  $T$ . More formally, for  $i$  ranging from 1 to  $n-1$ , let  $x$  be the largest leaf of  $T$ ; define  $w_i$  to be the unique neighbor of  $x$  in  $T$ ; then modify  $T$  by pruning the leaf  $x$ . This process produces a word  $w_1 \dots w_{n-1}$ ; we define  $f(T) = w_1 \dots w_{n-2}$ .



**FIGURE 3.16**

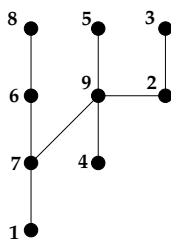
A tree with  $(d_1, \dots, d_9) = (1, 2, 1, 1, 1, 2, 3, 1, 4)$ .

**3.77. Example.** Let  $T$  be the tree shown in Figure 3.16. To compute  $f(T)$ , we prune leaves from  $T$  in the following order: 8, 5, 4, 9, 6, 3, 2, 7. Recording the neighbors of these leaves, we see that  $w = f(T) = 9996727$ . Observe that the algorithm computes  $w_{n-1} = 1$ , but this letter is not part of the output word  $w$ . Also observe that  $w \in \mathcal{R}(1^0 2^1 3^0 4^0 5^0 6^1 7^2 8^0 9^3) = \mathcal{R}(1^{d_1-1} \dots 9^{d_9-1})$ .

The observations in the last example hold in general. For, given any  $T \in A$ , repeatedly pruning leaves from  $T$  will produce a sequence of smaller trees, by the pruning lemma 3.69. By 3.67, each such tree (except the last tree) has at least two leaves, so vertex  $v_1$  will never be chosen for pruning. In particular,  $v_1$  is always the last vertex left, so that  $w_{n-1}$  is always  $v_1$ . Furthermore, if  $v_j$  is any vertex different from  $v_1$ , then the number of occurrences of  $v_j$  in  $w_1 w_2 \dots w_{n-1}$  is exactly  $d_j - 1$ . For, every time a pruning operation removes an edge touching  $v_j$ , we set  $w_i = v_j$  for some  $i$ , *except* when we are removing the last remaining edge touching  $v_j$  (which occurs when  $v_j$  has become the largest leaf and is being pruned). The same reasoning shows that  $v_1$  (which never gets pruned) appears  $d_1$  times in  $w_1 \dots w_{n-1}$ . Since  $w_{n-1} = v_1$ , *every* vertex  $v_j$  occurs  $d_j - 1$  times in the output word  $w_1 \dots w_{n-2}$ .

To see that  $f$  is a bijection, we argue by induction on the number of vertices. The result holds when  $n = 2$ , since in this case,  $A$  consists of a single tree with two nodes, and  $B$  consists of a single word (the empty word). Now suppose  $n > 2$  and the maps  $f$  (defined for trees with fewer than  $n$  vertices) are already known to be bijections. Given  $w = w_1 \dots w_{n-2} \in B$ , we will show there exists exactly one  $T \in A$  with  $f(T) = w$ . If such  $T$  exists, the leaves of  $T$  are precisely the vertices in  $V(T)$  that do *not* appear in  $w$ . Thus, the first leaf that gets pruned when computing  $f(T)$  must be the largest element  $z$  of  $V(T) \sim \{w_1, \dots, w_{n-2}\}$ . By induction hypothesis, there exists exactly one tree  $T'$  on the vertex set  $V(T) \sim \{z\}$  (with the appropriate vertex degrees) such that  $f(T') = w_2 \dots w_{n-2}$ . This given, we will have  $f(T) = w$  iff  $T$  is the tree obtained from  $T'$  by attaching a new leaf  $z$  as a neighbor of vertex  $w_1$ . One readily confirms that this graph *is* in  $A$  (i.e., the graph is a tree with the correct vertex degrees). This completes the induction argument. The proof also yields a recursive algorithm for computing  $f^{-1}(w)$ . The key point is to use the letters *not* seen in  $w$  (and its suffixes) to determine the identity of the leaf that was pruned at each stage.

**3.78. Example.** Given  $w = 6799297$  and  $V = \{1, 2, \dots, 9\}$ , let us compute the tree  $f^{-1}(w)$  with vertex set  $V$ . The leaves of this tree must be  $\{1, 3, 4, 5, 8\}$ , which are the elements of  $V$  not seen in  $w$ . Leaf 8 was pruned first and was adjacent to vertex 6. So now we must compute the tree  $f^{-1}(799297)$  with vertex set  $V \sim \{8\}$ . Here, leaf 6 was pruned first and was adjacent to vertex 7. Continuing in this way, we deduce that the leaves were pruned in the order 8, 6, 5, 4, 3, 2, 9, 7; and the neighbors of these leaves (reading from  $w$ ) were 6, 7, 9, 9, 2, 9, 7, 1. Thus,  $f^{-1}(w)$  is the tree shown in Figure 3.17.

**FIGURE 3.17**

The tree associated to  $w = 6799297$  by  $f^{-1}$ .

### 3.13 Ordered Trees and Terms

In this section, we study combinatorial structures called *ordered trees*, which are defined recursively as follows.

**3.79. Definition: Ordered Trees.** The symbol 0 is an ordered tree. If  $n > 0$  and  $T_1, \dots, T_n$  is a sequence of ordered trees, then the  $(n + 1)$ -tuple

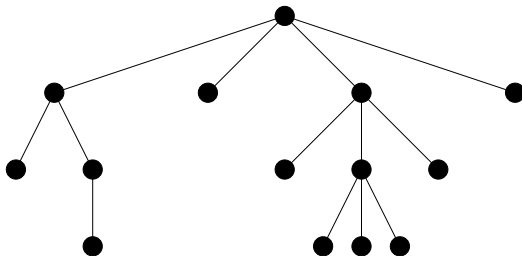
$$(n, T_1, T_2, \dots, T_n)$$

is an ordered tree. All ordered trees arise by applying these two rules a finite number of times.

We often think of ordered trees pictorially. The ordered tree 0 is depicted as a single node. The ordered tree  $(n, T_1, T_2, \dots, T_n)$  is drawn by putting a single “root node” at the top of the picture with  $n$  edges leading down. At the ends of these edges, reading from left to right, we recursively draw pictures of the trees  $T_1, T_2, \dots, T_n$  in this order. The term “ordered tree” emphasizes the fact that the left-to-right order of the children of each node is significant. Note that an ordered tree is *not* a tree in the graph-theoretic sense, and ordered trees are not the same as rooted trees.

**3.80. Example.** Figure 3.18 illustrates the ordered tree

$$T = (4, (2, 0, (1, 0)), 0, (3, 0, (3, 0, 0, 0), 0), 0).$$

**FIGURE 3.18**

Picture of an ordered tree.

Ordered trees can be used to model algebraic expressions that are built up by applying functions to lists of arguments. For example, the tree  $T$  in the previous example represents the syntactic structure of the following algebraic expression:

$$f(g(x_1, h(x_2)), x_3, k(x_4, j(x_5, x_6, x_7), x_8), x_9).$$

More specifically, if we replace each function symbol  $f, g, h, k, j$  by its *arity* (number of arguments) and replace each variable  $x_i$  by zero, we obtain

$$4(2(0, 1(0)), 0, 3(0, 3(0, 0, 0), 0), 0).$$

This becomes  $T$  if we move each left parenthesis to the left of the positive integer immediately preceding it and put a comma in its original location.

Surprisingly, it turns out that the syntactic structure of such an algebraic expression is uniquely determined even if we erase all the parentheses. To prove this statement, we introduce a combinatorial object called a *term* that is like an ordered tree, but contains no parentheses. For example, the algebraic expression above will be modeled by the term 42010030300000.

**3.81. Definition: Words and Terms.** A *word* is a finite sequence of natural numbers. We define *terms* recursively as follows. The word 0 is a term. If  $n > 0$  and  $T_1, T_2, \dots, T_n$  are terms, then the word  $nT_1T_2 \cdots T_n$  is a term. All terms arise by applying these two rules a finite number of times.

We see from this definition that every term is a *nonempty* word.

**3.82. Definition: Weight of a Word.** Given a word  $w = w_1w_2 \cdots w_s$ , the *weight* of  $w$  is  $\text{wt}(w) = w_1 + w_2 + \cdots + w_s - s$ .

For example,  $\text{wt}(42010030300000) = 13 - 14 = -1$ . Note that  $\text{wt}(vw) = \text{wt}(v) + \text{wt}(w)$  for all words  $v, w$ . The next result uses weights to characterize terms.

**3.83. Theorem: Characterization of Terms.** A word  $w = w_1w_2 \cdots w_s$  is a term iff  $\text{wt}(w) = -1$  and  $\text{wt}(w_1w_2 \cdots w_k) \geq 0$  for all  $k < s$ .

*Proof.* We argue by strong induction on the length  $s$  of the word. First suppose  $w = w_1w_2 \cdots w_s$  is a term of length  $s$ . If  $w = 0$ , then the weight condition holds. Otherwise, we must have  $w = nT_1T_2 \cdots T_n$  where  $n > 0$  and  $T_1, T_2, \dots, T_n$  are terms. Since each  $T_i$  has length less than  $s$ , the induction hypothesis shows that  $\text{wt}(T_i) = -1$  and every proper prefix of  $T_i$  has nonnegative weight. So, first of all,  $\text{wt}(w) = \text{wt}(n) + \text{wt}(T_1) + \cdots + \text{wt}(T_n) = (n - 1) - n = -1$ . On the other hand, consider a proper prefix  $w_1w_2 \cdots w_k$  of  $w$ . If  $k = 1$ , the weight of this prefix is  $n - 1$ , which is nonnegative since  $n > 0$ . If  $k > 1$ , we must have  $w_1w_2 \cdots w_k = nT_1 \cdots T_i z$  where  $0 \leq i < n$  and  $z$  is a proper prefix of  $T_{i+1}$ . Using the induction hypothesis, the weight of  $w_1w_2 \cdots w_k$  is therefore  $(n-1) - i + \text{wt}(z) \geq (n-i) - 1 \geq 0$ .

For the converse, we also use strong induction on the length of the word. Let  $w = w_1w_2 \cdots w_s$  satisfy the weight conditions. The empty word has weight zero, so  $s > 0$ . If  $s = 1$ , then  $\text{wt}(w_1) = -1$  forces  $w = 0$ , so that  $w$  is a term in this case. Now suppose  $s > 1$ . The first symbol  $w_1$  must be an integer  $n > 0$ , lest the proper prefix  $w_1$  of  $w$  have negative weight. Observe that appending one more letter to any word decreases the weight by at most 1. Since  $\text{wt}(w_1) = n - 1$  and  $\text{wt}(w_1w_2 \cdots w_s) = -1$ , there exists a least integer  $k_1$  with  $\text{wt}(w_1w_2 \cdots w_{k_1}) = n - 2$ . Now if  $n \geq 2$ , there exists a least integer  $k_2 > k_1$  with  $\text{wt}(w_1w_2 \cdots w_{k_2}) = n - 3$ . We continue similarly, obtaining integers  $k_1 < k_2 < \cdots < k_n$  such that  $k_i$  is the least index following  $k_{i-1}$  such that  $\text{wt}(w_1w_2 \cdots w_{k_i}) = n - 1 - i$ . Because  $w$  satisfies the weight conditions, we must have  $k_n = s$ . Now define  $n$  subwords

$T_1 = w_2 w_3 \cdots w_{k_1}$ ,  $T_2 = w_{k_1+1} w_{k_1+2} \cdots w_{k_2}$ ,  $\dots$ ,  $T_n = w_{k_{n-1}+1} \cdots w_s$ . Evidently  $w = nT_1 T_2 \cdots T_n$ . For all  $i \leq n$ ,  $nT_1 T_2 \cdots T_{i-1}$  has weight  $n-i$ ,  $nT_1 T_2 \cdots T_i$  has weight  $n-i-1$ , and (by minimality of  $k_i$ ) no proper prefix of  $nT_1 T_2 \cdots T_i$  has weight less than  $n-i$ . It follows that  $T_i$  has weight  $-1$  but every proper prefix of  $T_i$  has nonnegative weight. Thus each  $T_i$  satisfies the weight conditions and has length less than  $w$ . By induction, every  $T_i$  is a term. Then  $w = nT_1 T_2 \cdots T_n$  is also a term, completing the induction.  $\square$

**3.84. Corollary.** No proper prefix of a term is a term.

**3.85. Theorem: Unique Readability of Terms.** For every term  $w$ , there exists a unique integer  $n \geq 0$  and unique terms  $T_1, \dots, T_n$  such that  $w = nT_1 \cdots T_n$ .

*Proof.* Existence follows from the recursive definition of terms. We prove uniqueness by induction on the length of  $w$ . Suppose  $w = nT_1 \cdots T_n = mT'_1 \cdots T'_m$  where  $n, m \geq 0$  and every  $T_i$  and  $T'_j$  is a term. We must prove  $n = m$  and  $T_i = T'_i$  for all  $i \leq n$ . First,  $n = w_1 = m$ . If  $T_1 \neq T'_1$ , then one of  $T_1$  and  $T'_1$  must be a proper prefix of the other, in violation of the preceding corollary. So  $T_1 = T'_1$ . Then if  $T_2 \neq T'_2$ , one of  $T_2$  and  $T'_2$  must be a proper prefix of the other, in violation of the corollary. Continuing similarly, we see that  $T_i = T'_i$  for all  $i$ .  $\square$

Using the previous theorem and induction, one readily proves that erasing all parentheses defines a bijection from ordered trees to terms. Therefore, to enumerate various collections of ordered trees, it suffices to enumerate the corresponding collections of terms. This technique will be used in the next section.

## 3.14 Ordered Forests and Lists of Terms

We continue our study of ordered trees and terms by introducing two more general objects: ordered forests and lists of terms.

**3.86. Definition: Ordered Forests.** For  $n \geq 0$ , an *ordered forest of  $n$  trees* is a list  $(T_1, T_2, \dots, T_n)$ , where each  $T_i$  is an ordered tree.

**3.87. Definition: Lists of Terms.** For  $n \geq 0$ , a *list of  $n$  terms* is a word  $w$  of the form  $w = T_1 T_2 \cdots T_n$ , where each  $T_i$  is a term.

**3.88. Theorem: Weight Characterization of Lists of Terms.** A word  $w = w_1 w_2 \cdots w_s$  is a list of  $n$  terms iff  $\text{wt}(w) = -n$  and  $\text{wt}(w_1 w_2 \cdots w_k) > -n$  for all  $k < s$ .

*Proof.* First suppose  $w$  is a list of  $n$  terms, say  $w = T_1 T_2 \cdots T_n$ . Then  $nw = nT_1 T_2 \cdots T_n$  is a single term. This term has weight  $-1$ , by 3.83, so  $w$  has weight  $-1 - \text{wt}(n) = -n$ . If  $\text{wt}(w_1 \cdots w_k) \leq -n$  for some  $k < s$ , then the proper prefix  $nw_1 \cdots w_k$  of the term  $nw$  would have negative weight, contradicting 3.83.

Conversely, suppose  $w$  satisfies the weight conditions in 3.88. Then the word  $nw$  satisfies the weight conditions in 3.83, as one immediately verifies. So  $nw$  is a term, which must have the form  $nT_1 T_2 \cdots T_n$  for suitable terms  $T_1, \dots, T_n$ . Then  $w = T_1 T_2 \cdots T_n$  is a list of  $n$  terms.  $\square$

**3.89. Theorem: Unique Readability of Lists of Terms.** If  $w = T_1 T_2 \cdots T_n$  is a list of  $n$  terms, then  $n$  and the terms  $T_i$  are uniquely determined by  $w$ .

*Proof.* First,  $n = -\text{wt}(w)$  is uniquely determined by  $w$ . To see that the  $T_i$  are unique, add an  $n$  to the beginning of  $w$  and then appeal to 3.85.  $\square$

We deduce that erasing parentheses gives a bijection between ordered forests of  $n$  trees and lists of  $n$  terms.

The next lemma reveals a key property that will allow us to enumerate lists of terms.

**3.90. Cycle Lemma for Lists of Terms.** Suppose  $w = w_1 w_2 \cdots w_s$  is a word of weight  $-n < 0$ . There exist exactly  $n$  indices  $i \leq s$  such that the cyclic rotation

$$R_i(w) = w_i w_{i+1} \cdots w_s w_1 w_2 \cdots w_{i-1}$$

is a list of  $n$  terms.

*Proof. Step 1:* We prove the result when  $w$  itself is a list of  $n$  terms. Say  $w = T_1 T_2 \cdots T_n$  where  $T_j$  is a term of length  $k_j$ . Then  $R_i(w)$  is a list of  $n$  terms for the  $n$  indices  $i \in \{1, k_1 + 1, k_1 + k_2 + 1, \dots, k_1 + k_2 + \cdots + k_{n-1} + 1\}$ . Suppose  $i$  is another index (different from those just listed) such that  $R_i(w)$  is a list of  $n$  terms. For some  $j \leq n$ , we must have

$$R_i(w) = y T_{j+1} \cdots T_n T_1 \cdots T_{j-1} z$$

where  $T_j = zy$  and  $z, y$  are nonempty words. Since  $\text{wt}(z) \geq 0$  but  $\text{wt}(T_j) = -1$ , we must have  $\text{wt}(y) < 0$ . So

$$\text{wt}(y T_{j+1} \cdots T_{j-1}) = \text{wt}(y) + \text{wt}(T_{j+1}) + \cdots + \text{wt}(T_{j-1}) < -(n-1).$$

Then  $y T_{j+1} \cdots T_{j-1}$  is a proper prefix of  $R_i(w)$  with weight  $\leq -n$ , in violation of 3.88.

*Step 2:* We prove the result for a general word  $w$ . It suffices to show that there exists at least one  $i$  such that  $R_i(w)$  is a list of  $n$  terms. For then, since we obtain the same collection of words by cyclically shifting  $w$  and  $R_i(w)$ , the desired result will follow from Step 1.

First note that all cyclic rotations of  $w$  have weight  $\sum_{j=1}^s \text{wt}(w_j) = \text{wt}(w) = -n$ . Let  $m$  be the minimum weight of any prefix  $w_1 w_2 \cdots w_k$  of  $w$ , where  $1 \leq k \leq s$ . Choose  $k$  minimal such that  $\text{wt}(w_1 w_2 \cdots w_k) = m$ . If  $k = s$ , then  $m = -n$ , and by minimality of  $k$  and 3.88,  $w$  itself is already a list of  $n$  terms. Otherwise, let  $i = k + 1$ . We claim  $R_i(w)$  is a list of  $n$  terms. It suffices to check that each proper prefix of  $R_i(w)$  has weight  $> -n$ . On one hand, for all  $j$  with  $i \leq j \leq s$ , the prefix  $w_i \cdots w_j$  of  $R_i(w)$  cannot have negative weight; otherwise,  $\text{wt}(w_1 \cdots w_k w_i \cdots w_j) < m$  violates the minimality of  $m$ . So  $\text{wt}(w_i \cdots w_j) \geq 0 > -n$ . Note that when  $j = s$ , we have  $\text{wt}(w_i \cdots w_s) = \text{wt}(w) - \text{wt}(w_1 \cdots w_k) = -n - m$ . Now consider  $j$  in the range  $1 \leq j < k$ . If  $\text{wt}(w_i \cdots w_s w_1 \cdots w_j) \leq -n$ , then

$$\text{wt}(w_1 \cdots w_j) = \text{wt}(w_i \cdots w_s w_1 \cdots w_j) - \text{wt}(w_i \cdots w_s) \leq -n - (-n - m) = m.$$

But this violates the choice of  $k$  as the *least* index such that the prefix ending at  $k$  has minimum weight. So  $\text{wt}(w_i \cdots w_s w_1 \cdots w_j) > -n$ . It now follows from 3.88 that  $R_i(w)$  is indeed a list of  $n$  terms.  $\square$

Suppose  $w$  is a list of  $n$  terms containing exactly  $k_i$  occurrences of  $i$  for each  $i \geq 0$ . We have

$$-n = \text{wt}(w) = \sum_{i \geq 0} k_i \text{wt}(i) = \sum_{i \geq 0} (i-1)k_i = -k_0 + \sum_{i \geq 1} (i-1)k_i.$$

It follows that  $k_0 = n + \sum_{i \geq 1} (i-1)k_i$  in this situation. Conversely, if  $k_0$  satisfies this relation, then  $\text{wt}(w) = -n$  for all  $w \in \mathcal{R}(0^{k_0} 1^{k_1} 2^{k_2} \cdots)$ . We now have all the ingredients needed for our main enumeration result.

**3.91. Theorem: Enumeration of Lists of Terms.** Let  $n > 0$  and  $k_0, k_1, \dots, k_t$  be given natural numbers such that  $k_0 = n + \sum_{i=1}^t (i-1)k_i$ . The number  $N$  of words  $w$  such that  $w$  is a list of  $n$  terms containing  $k_i$  copies of  $i$  for  $0 \leq i \leq t$  is

$$\frac{n}{s} \binom{s}{k_0, k_1, \dots, k_t} = \frac{n(s-1)!}{k_0!k_1!\cdots k_t!},$$

where  $s = \sum_{i=0}^t k_i = n + \sum_{i=1}^t ik_i$  is the common length of all such words.

*Proof. Step 1.* Let  $A$  be the set of all pairs  $(w, j)$ , where  $w \in \mathcal{R}(0^{k_0}1^{k_1}\cdots t^{k_t})$  is a word and  $j \leq s$  is an index such that the cyclic rotation  $R_j(w)$  is a list of  $n$  terms. Combining 3.90 and 1.46, we see that  $|A| = n \binom{s}{k_0, k_1, \dots, k_t}$ .

*Step 2.* Let  $B$  be the set of all pairs  $(w, i)$  where  $w \in \mathcal{R}(0^{k_0}1^{k_1}\cdots t^{k_t})$  is a list of  $n$  terms (necessarily of length  $s$ ) and  $1 \leq i \leq s$ . By the product rule,  $|B| = sN$ .

*Step 3.* To complete the proof, we show that  $|A| = |B|$  by exhibiting mutually inverse bijections  $f: A \rightarrow B$  and  $g: B \rightarrow A$ . We define  $f(w, j) = (R_j(w), j)$  for all  $(w, j) \in A$ , and  $g(w, i) = (R_i^{-1}(w), i)$  for all  $(w, i) \in B$ .  $\square$

## 3.15 Graph Coloring

This section introduces the graph coloring problem and some of its applications.

**3.92. Definition: Colorings.** Let  $G = (V, E)$  be a simple graph, and let  $C$  be a finite set. A *coloring* of  $G$  using colors in  $C$  is a function  $f: V \rightarrow C$ . A coloring  $f$  of  $G$  is a *proper coloring* iff for every edge  $\{u, v\} \in E$ ,  $f(u) \neq f(v)$ .

Intuitively, we are coloring each vertex of  $G$  using one of the available colors in the set  $C$ . For each  $v \in V$ ,  $f(v)$  is the color assigned to vertex  $v$ . A coloring is proper iff no two adjacent vertices in  $G$  are assigned the same color.

**3.93. Definition: Chromatic Functions and Chromatic Numbers.** Let  $G$  be a simple graph. For each positive integer  $x$ , let  $\chi_G(x)$  be the number of proper colorings of  $G$  using colors in  $\{1, 2, \dots, x\}$ . The function  $\chi_G: \mathbb{N}^+ \rightarrow \mathbb{N}$  is called the *chromatic function* of  $G$ . The minimal  $x$  such that  $\chi_G(x) > 0$  is called the *chromatic number* of  $G$ .

The chromatic number is the least number of colors required to obtain a proper coloring of  $G$ . The function  $\chi_G$  is often called the *chromatic polynomial* of  $G$  because of 3.100 below.

**3.94. Example.** Suppose  $G$  is a simple graph with  $n$  vertices and no edges. Then  $\chi_G(x) = x^n$  since we can assign any of the  $x$  colors to each vertex. The chromatic number for this graph is 1.

**3.95. Example.** At the other extreme, suppose  $G$  is a simple graph with  $n$  vertices such that there is an edge joining every pair of distinct vertices. Color the vertices one at a time. The first vertex can be colored in  $x$  ways. The second vertex must have a color different from the first, so there are  $x-1$  choices. In general, the  $i$ th vertex must have a color distinct from all of its predecessors, so there are  $x-(i-1)$  choices for the color of this vertex. The product rule gives  $\chi_G(x) = x(x-1)(x-2)\cdots(x-n+1) = (x)_{\downarrow n}$ . The chromatic number for this graph is  $n$ . Recall from §2.13 that

$$(x)_{\downarrow n} = \sum_{k=1}^n s(n, k)x^k,$$

so that the function  $\chi_G$  in this example is a polynomial whose coefficients are the signed Stirling numbers of the first kind.

**3.96. Example: Cycles.** Consider the simple graph

$$G = (\{1, 2, 3, 4\}, \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 1\}\}).$$

$G$  consists of four vertices joined in a 4-cycle. We might attempt to compute  $\chi_G(x)$  via the product rule as follows. Color vertex 1 in  $x$  ways. Then color vertex 2 in  $x-1$  ways, and color vertex 3 in  $x-1$  ways. We run into trouble at vertex 4, because we do not know whether vertices 1 and 3 were assigned the same color. This example shows that we cannot always compute  $\chi_G$  by the product rule alone. In this instance, we can classify proper colorings based on whether vertices 1 and 3 receive the same or different colors. If they receive the same color, the number of proper colorings is  $x(x-1)(x-1)$  (color vertices 1 and 3 together, then color vertex 2 a different color, then color vertex 4 a different color from 1 and 3). If vertex 1 and 3 receive different colors, the number of proper colorings is  $x(x-1)(x-2)(x-2)$  (color vertex 1, then vertex 3, then vertex 2, then vertex 4). Hence

$$\chi_G(x) = x(x-1)(x-1) + x(x-1)(x-2)(x-2) = x^4 - 4x^3 + 6x^2 - 3x.$$

The chromatic number for this graph is 2.

More generally, consider the graph  $C_n$  consisting of  $n$  vertices joined in a cycle. It is routine to establish that the chromatic number of  $C_n$  is 1 for  $n = 1$ , is 2 for all even  $n$ , and is 3 for all odd  $n > 1$ . On the other hand, it is not immediately evident how to compute the chromatic function for  $C_n$  when  $n > 4$ . We will deduce a recursion for these functions shortly as a special case of a general recursion for computing chromatic functions.

Here is an application that can be analyzed using graph colorings and chromatic numbers. Suppose we are trying to schedule meetings for a number of committees. If two committees share a common member, they cannot meet at the same time. Consider the graph  $G$  whose vertices represent the various committees, and where there is an edge between two vertices iff the corresponding committees share a common member. Suppose there are  $x$  available time slots in which meetings may be scheduled. A coloring of  $G$  with  $x$  colors represents a particular scheduling of committee meetings to time slots. The coloring is proper iff the schedule creates no time conflicts for any committee member. The chromatic number is the least number of time slots needed to avoid all conflicts, while  $\chi_G(x)$  is the number of different conflict-free schedules using  $x$  (distinguishable) time slots.

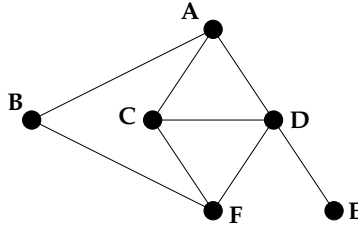
**3.97. Example.** Six committees have members as specified in the following table.

Committee	Members
A	Kemp, Oakley, Saunders
B	Gray, Saunders, Russell
C	Byrd, Oakley, Quinn
D	Byrd, Jenkins, Kemp
E	Adams, Jenkins, Wilson
F	Byrd, Gray, Russell

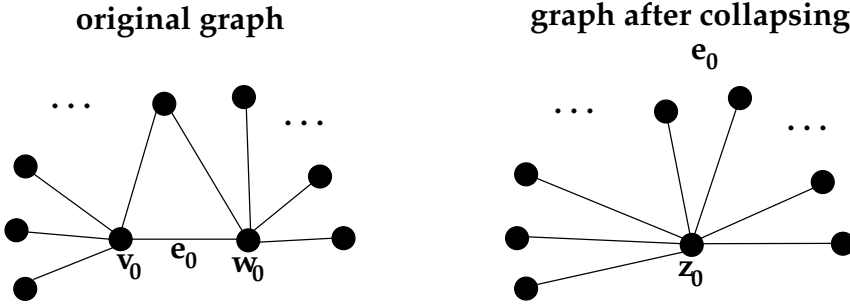
Figure 3.19 displays the graph  $G$  associated to this set of committees. To compute  $\chi_G(x)$ , consider cases based on whether vertices A and F receive the same color. If A and F are colored the same, the number of proper colorings is  $x(x-1)(x-2)(x-1)(x-1)$  [color A and F, then C, D, B, and E]. If A and F receive different colors, the number of proper colorings is  $x(x-1)(x-2)(x-3)(x-2)(x-1)$  [color A, F, C, D, B, E]. Thus,

$$\chi_G(x) = x(x-1)^2(x-2)(x-1 + (x-2)(x-3)) = x^6 - 8x^5 + 26x^4 - 42x^3 + 33x^2 - 10x.$$

The chromatic number of  $G$  is 3.

**FIGURE 3.19**

Conflict graph for six committees.

**FIGURE 3.20**

Collapsing an edge in a simple graph.

We are about to present a general recursion that can be used to compute the chromatic function of a simple graph. The recursion makes use of the following construction.

**3.98. Definition: Collapsing an Edge.** Let  $G = (V, E)$  be a simple graph, and let  $e_0 = \{v_0, w_0\}$  be a fixed edge of  $G$ . Let  $z_0$  be a new vertex. We define a simple graph  $H$  called *the graph obtained from  $G$  by collapsing the edge  $e_0$* . The vertex set of  $H$  is  $(V \sim \{v_0, w_0\}) \cup \{z_0\}$ . The edge set of  $H$  is

$$\begin{aligned} & \{\{x, y\} : x \neq v_0 \neq y \text{ and } x \neq w_0 \neq y \text{ and } \{x, y\} \in E\} \\ & \cup \{\{x, z_0\} : x \neq v_0 \text{ and } \{x, w_0\} \in E\} \\ & \cup \{\{x, z_0\} : x \neq w_0 \text{ and } \{x, v_0\} \in E\}. \end{aligned}$$

Pictorially, we construct  $H$  from  $G$  by shrinking the edge  $e_0$  until the vertices  $v_0$  and  $w_0$  coincide. We replace these two overlapping vertices with a single new vertex  $z_0$ . All edges touching  $v_0$  or  $w_0$  (except the collapsed edge  $e_0$ ) now touch  $z_0$  instead. See Figure 3.20.

**3.99. Theorem: Chromatic Recursion.** Let  $G = (V, E)$  be a simple graph. Fix any edge  $e = \{v, w\} \in G$ . Let  $G' = (V, E \sim \{e\})$  be the simple graph obtained by deleting the edge  $e$  from  $G$ , and let  $G''$  be the simple graph obtained from  $G$  by collapsing the edge  $e$ . Then

$$\chi_G(x) = \chi_{G'}(x) - \chi_{G''}(x).$$

*Proof.* Fix  $x \in \mathbb{N}^+$ , and let  $A$ ,  $B$ , and  $C$  denote the set of proper colorings of  $G$ ,  $G'$ , and  $G''$  (respectively) using  $x$  available colors. Write  $B = B_1 \cup B_2$ , where  $B_1 = \{f \in B : f(v) = f(w)\}$  and  $B_2 = \{f \in B : f(v) \neq f(w)\}$ . Note that  $B_1$  consists of the proper colorings of  $G'$  (if any) in which vertices  $v$  and  $w$  are assigned the same color. Let  $z$  be the new vertex



in  $G''$  that replaces  $v$  and  $w$ . Given a proper coloring  $f \in B_1$ , we define a corresponding coloring  $f''$  of  $G''$  by setting  $f''(z) = f(v) = f(w)$  and  $f''(u) = f(u)$  for all  $u \in V$  different from  $v$  and  $w$ . Since  $f$  is proper, it follows from the definition of the edge set of  $G''$  that  $f''$  is a proper coloring as well. Thus we have a map  $f \mapsto f''$  from  $B_1$  to  $C$ . This map is invertible, since the color of  $z$  in a coloring of  $G''$  determines the common color of  $v$  and  $w$  in a coloring of  $G'$  belonging to  $B_1$ . We conclude that  $|B_1| = |C|$ .

On the other hand,  $B_2$  consists of the proper colorings of  $G'$  in which vertices  $v$  and  $w$  are assigned different colors. These are precisely the proper colorings of  $G$  (since  $G$  has an edge between  $v$  and  $w$ , and  $G$  is otherwise identical to  $G'$ ). Thus,  $B_2 = A$ . It follows that

$$\chi_G(x) = |A| = |B_2| = |B| - |B_1| = |B| - |C| = \chi_{G'}(x) - \chi_{G''}(x). \quad \square$$

**3.100. Corollary: Polynomiality of Chromatic Functions.** For any graph  $G$ ,  $\chi_G(x)$  is a polynomial in  $x$  with integer coefficients. (This justifies the terminology *chromatic polynomial*.)

*Proof.* We use induction on the number of edges in  $G$ . If  $G$  has  $k$  vertices and no edges, the product rule gives  $\chi_G(x) = x^k$ , which is a polynomial in  $x$ . Now assume  $G$  has  $m > 0$  edges. Fix such an edge  $e$ , and define  $G'$  and  $G''$  as in the preceding theorem.  $G'$  has one fewer edge than  $G$ . When passing from  $G$  to  $G''$ , we lose the edge  $e$  and possibly identify other edges in  $G$  (e.g., if both endpoints of  $e$  are adjacent to a third vertex). In any case,  $G''$  has fewer edges than  $G$ . By induction on  $m$ , we may assume that both  $\chi_{G'}(x)$  and  $\chi_{G''}(x)$  are polynomials in  $x$  with integer coefficients. So  $\chi_G(x) = \chi_{G'}(x) - \chi_{G''}(x)$  is also a polynomial with integer coefficients.  $\square$

**3.101. Remark.** We can use the chromatic recursion to compute  $\chi_G$  recursively for any graph  $G$ . The base case of the calculation is a graph with  $k$  vertices and no edges, which has chromatic polynomial  $x^k$ . If  $G$  has more than one edge,  $G'$  and  $G''$  both have strictly fewer edges than  $G$ . Thus, the recursive calculation will terminate after finitely many steps. However, this is quite an inefficient method for computing  $\chi_G$  if  $G$  has many vertices and edges. Thus, direct counting arguments using the sum and product rules may be preferable to repeatedly applying the chromatic recursion.

**3.102. Example.** Consider the graph  $G$  shown on the left in Figure 3.21. We compute  $\chi_G(x)$  by applying the chromatic recursion to the edge  $e = \{d, h\}$ . The graphs  $G'$  and  $G''$  obtained by deleting and collapsing this edge are shown on the right in Figure 3.21. Direct arguments using the product rule show that

$$\chi_{G'}(x) = x(x-1)(x-2)(x-2)(x-1)(x-1) \quad (\text{color } a, c, d, f, b, h);$$

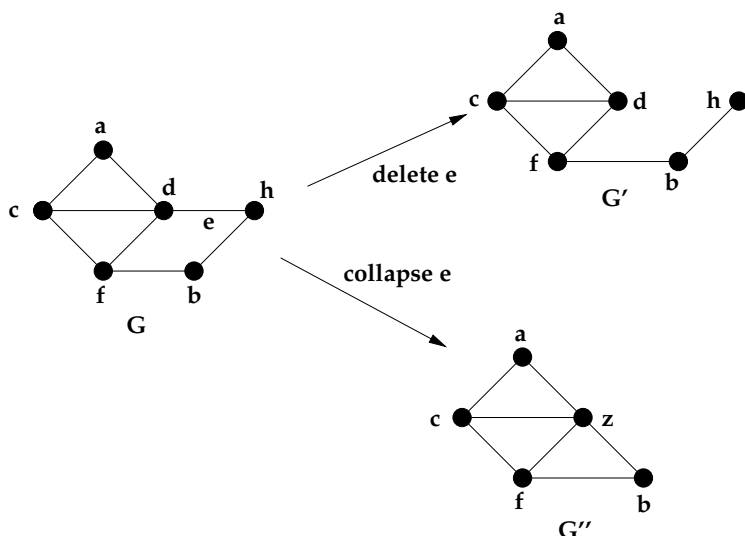
$$\chi_{G''}(x) = x(x-1)(x-2)(x-2)(x-2) \quad (\text{color } z, a, c, f, b).$$

Therefore,

$$\chi_G(x) = x(x-1)(x-2)^2((x-1)^2 - (x-2)) = x^6 - 8x^5 + 26x^4 - 43x^3 + 36x^2 - 12x.$$

**3.103. Chromatic Polynomials for Cycles.** For each  $n \geq 3$ , let  $C_n$  denote a graph consisting of  $n$  vertices joined in a cycle. Let  $C_1$  denote a one-vertex graph, and let  $C_2$  denote a graph with two vertices joined by an edge. Finally, let  $\chi_n(x) = \chi_{C_n}(x)$  be the chromatic polynomials for these graphs. We see directly that

$$\chi_1(x) = x, \quad \chi_2(x) = x(x-1) = x^2 - x, \quad \chi_3(x) = x(x-1)(x-2) = x^3 - 3x^2 + 2x.$$


**FIGURE 3.21**

Using the chromatic recursion.

Fix  $n > 3$  and fix any edge  $e$  in  $C_n$ . Deleting this edge leaves a graph in which  $n$  vertices are joined in a line; the chromatic polynomial of such a graph is  $x(x-1)^{n-1}$ . On the other hand, collapsing the edge  $e$  in  $C_n$  produces a graph isomorphic to  $C_{n-1}$ . The chromatic recursion therefore gives

$$\chi_n(x) = x(x-1)^{n-1} - \chi_{n-1}(x).$$

Using this recursion to compute  $\chi_n(x)$  for small  $n$  suggests the closed formula

$$\chi_n(x) = (x-1)^n + (-1)^n(x-1) \quad (n \geq 2).$$

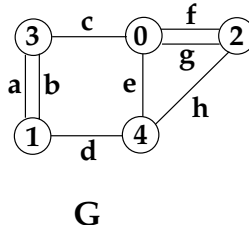
We let the reader prove this formula for  $\chi_n(x)$  by induction, using the chromatic recursion.

### 3.16 Spanning Trees

This section introduces the notion of a spanning tree for a graph. A recursion resembling the chromatic recursion 3.99 will allow us to count the spanning trees for a given graph. This will lead to a remarkable formula, called the matrix-tree theorem, that expresses the number of spanning trees as a certain determinant.

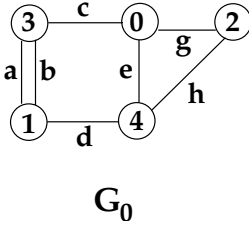
**3.104. Definition: Subgraphs.** Let  $G = (V, E, \epsilon)$  and  $H = (W, F, \eta)$  be graphs or digraphs.  $H$  is a *subgraph* of  $G$  iff  $W \subseteq V$ ,  $F \subseteq E$ , and  $\eta(f) = \epsilon(f)$  for all  $f \in F$ .  $H$  is an *induced subgraph* of  $G$  iff  $H$  is a subgraph such that  $F$  consists of *all* edges in  $E$  with both endpoints in  $W$ .

**3.105. Definition: Spanning Trees.** Given a graph  $G = (V, E, \epsilon)$ , a *spanning tree* for  $G$  is a subgraph  $H$  with vertex set  $V$  such that  $H$  is a tree. Let  $\tau(G)$  be the number of spanning trees of  $G$ .

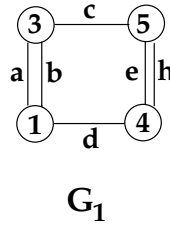
**FIGURE 3.22**

Graph used to illustrate spanning trees.

**delete edge f:**



**collapse edge f:**

**FIGURE 3.23**

Effect of deleting or collapsing an edge.

**3.106. Example.** Consider the graph  $G$  shown in Figure 3.22. This graph has 31 spanning trees, which are specified by the following sets of edges:

$$\begin{aligned}
 &\{a, c, d, f\}, \quad \{b, c, d, f\}, \quad \{a, c, d, g\}, \quad \{b, c, d, g\}, \quad \{a, c, d, h\}, \quad \{b, c, d, h\}, \\
 &\{c, d, e, f\}, \quad \{c, d, e, g\}, \quad \{c, d, e, h\}, \quad \{a, c, e, f\}, \quad \{a, d, e, f\}, \quad \{b, c, e, f\}, \\
 &\{b, d, e, f\}, \quad \{a, c, e, g\}, \quad \{a, d, e, g\}, \quad \{b, c, e, g\}, \quad \{b, d, e, g\}, \quad \{a, c, e, h\}, \\
 &\{a, d, e, h\}, \quad \{b, c, e, h\}, \quad \{b, d, e, h\}, \quad \{a, c, f, h\}, \quad \{a, d, f, h\}, \quad \{b, c, f, h\}, \\
 &\{b, d, f, h\}, \quad \{a, c, g, h\}, \quad \{a, d, g, h\}, \quad \{b, c, g, h\}, \quad \{b, d, g, h\}, \quad \{c, d, f, h\}, \\
 &\{c, d, g, h\}.
 \end{aligned}$$

We see that even a small graph can have many spanning trees. Thus we seek a systematic method for enumerating these trees.

We are going to derive a recursion involving the quantities  $\tau(G)$ . For this purpose, we need to adapt the ideas of *deleting an edge* and *collapsing an edge* (see 3.98) from simple graphs to general graphs. Since loop edges are never involved in spanning trees, we will only consider graphs without loops. Suppose we are given a graph  $G = (V, E, \epsilon)$  and a fixed edge  $z \in E$  with endpoints  $u, v \in V$ . To *delete*  $z$  from  $G$ , we replace  $E$  by  $E' = E \sim \{z\}$  and replace  $\epsilon$  by the restriction of  $\epsilon$  to  $E'$ . To *collapse* the edge  $z$ , we act as follows: (i) delete  $z$  and any other edges linking  $u$  to  $v$ ; (ii) replace  $V$  by  $(V \sim \{u, v\}) \cup \{w\}$ , where  $w$  is a new vertex; (iii) for each edge  $y \in E$  that has exactly one endpoint in the set  $\{u, v\}$ , modify  $\epsilon(y)$  by replacing this endpoint with the new vertex  $w$ .

**3.107. Example.** Let  $G$  be the graph shown in Figure 3.22. Figure 3.23 displays the graphs obtained from  $G$  by deleting edge  $f$  and collapsing edge  $f$ .

**3.108. Theorem: Spanning Tree Recursion.** Let  $G = (V, E, \epsilon)$  be a graph, and let  $z \in E$  be a fixed edge. Let  $G_0$  be the graph obtained from  $G$  by deleting  $z$ . Let  $G_1$  be the graph obtained from  $G$  by collapsing  $z$ . Then

$$\tau(G) = \tau(G_0) + \tau(G_1).$$

The initial conditions are:  $\tau(G) = 0$  if  $G$  is not connected, and  $\tau(G) = 1$  if  $G$  is a tree with vertex set  $V$ .

*Proof.* For every graph  $K$ , let  $Sp(K)$  be the set of all spanning trees of  $K$ , so  $\tau(K) = |Sp(K)|$ . Fix the graph  $G$  and the edge  $z$ . Let  $X$  be the set of trees in  $Sp(G)$  that do not use the edge  $z$ , and let  $Y$  be the set of trees in  $Sp(G)$  that do use the edge  $z$ .  $Sp(G)$  is the disjoint union of  $X$  and  $Y$ , so  $\tau(G) = |X| + |Y|$  by the sum rule. Now, it follows from the definition of edge-deletion that the set  $X$  is precisely the set  $Sp(G_0)$ , so  $|X| = \tau(G_0)$ . To complete the proof, we need to show that  $|Y| = \tau(G_1)$ . It suffices to define a bijection  $F : Y \rightarrow Sp(G_1)$ .

Suppose  $T \in Y$  is a spanning tree of  $G$  that uses the edge  $z$  with endpoints  $u, v$ . Define  $F(T)$  to be the graph obtained from  $T$  by collapsing the edge  $z$ ; this graph is a subgraph of  $G_1$ . Let  $n$  be the number of vertices of  $G$ ; then  $T$  is a connected graph with  $n - 1$  edges, one of which is  $z$ . It is routine to check that  $F(T)$  is still connected. Furthermore, since  $T$  is a tree,  $z$  is the only edge in  $T$  between  $u$  and  $v$ . It follows from the definition of collapsing that  $F(T)$  has exactly  $n - 2$  edges. Since  $G_1$  has  $n - 1$  vertices, it follows that  $F(T)$  is a spanning tree of  $G_1$ . We see also that the edge set of  $F(T)$  is precisely the edge set of  $T$  with  $z$  removed. So far, we have shown that  $F$  is a well-defined function mapping  $Y$  into  $Sp(G_1)$ .

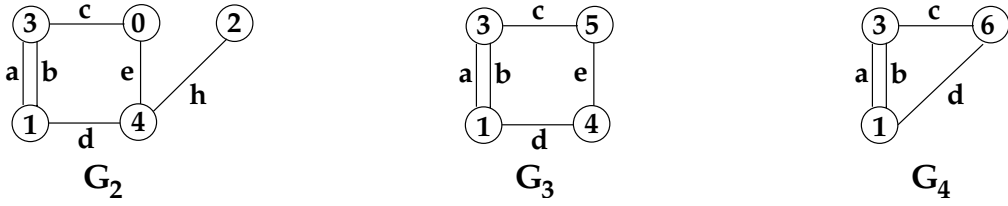
Next we define a map  $H : Sp(G_1) \rightarrow Y$  that will be the two-sided inverse of  $F$ . Given  $U \in Sp(G_1)$  with edge set  $E(U)$ , let  $H(U)$  be the unique subgraph of  $G$  with vertex set  $V$  and edge set  $E(U) \cup \{z\}$ . We must check that  $H(U)$  does lie in the claimed codomain  $Y$ . First,  $H(U)$  is a subgraph of  $G$  with  $n - 1$  edges, one of which is the edge  $z$ . Furthermore, one may check that  $H(U)$  is connected, since walks in  $U$  can be expanded using the edge  $z$  if needed to give walks in  $H(U)$ . Therefore,  $H(U)$  is a spanning tree of  $G$  using  $z$ , and so  $H(U) \in Y$ . Since  $F$  removes  $z$  from the edge set while  $H$  adds it back,  $F$  and  $H$  are two-sided inverses of each other. Hence both are bijections, and the proof is complete.  $\square$

**3.109. Example.** Let us use the graphs in Figures 3.22 and 3.23 to illustrate the proof of the spanning tree recursion, taking  $z = f$ . The graph  $G_0$  on the left of Figure 3.23 has 19 spanning trees; they are precisely the trees listed in 3.106 that do not use the edge  $f$ . Applying  $F$  to each of the remaining 12 spanning trees on the list produces the following subgraphs of  $G_1$  (specified by their edge sets):

$$\begin{array}{cccccc} \{a, c, d\}, & \{b, c, d\}, & \{c, d, e\}, & \{a, c, e\}, & \{a, d, e\}, & \{b, c, e\}, \\ \{b, d, e\}, & \{a, c, h\}, & \{a, d, h\}, & \{b, c, h\}, & \{b, d, h\}, & \{c, d, h\}. \end{array}$$

These are precisely the spanning trees of  $G_1$ .

Next, we illustrate the calculation of  $\tau(G)$  using the recursion. We first delete and collapse edge  $f$ , producing the graphs  $G_0$  and  $G_1$  shown in Figure 3.23. We know that  $\tau(G) = \tau(G_0) + \tau(G_1)$ . Deletion of edge  $g$  from  $G_0$  produces a new graph  $G_2$  (Figure 3.24), while collapsing  $g$  in  $G_0$  leads to another copy of  $G_1$ . So far, we have  $\tau(G) = 2\tau(G_1) + \tau(G_2)$ . Continuing to work on  $G_1$ , we see that deleting (resp. collapsing) edge  $h$  leads to the graph  $G_3$  (resp.  $G_4$ ) in Figure 3.24. On the other hand, deleting  $h$  from  $G_2$  leaves a disconnected graph (which can be discarded), while collapsing  $h$  from  $G_2$  produces another copy of  $G_3$ . Now we have  $\tau(G) = 3\tau(G_3) + 2\tau(G_4)$ . Deleting edge  $e$  from  $G_3$  gives a graph that has two spanning trees (by inspection), while collapsing  $e$  in  $G_3$  leads to  $G_4$  again. So

**FIGURE 3.24**

Auxiliary graphs used in the computation of  $\tau(G)$ .

$\tau(G) = 3(2 + \tau(G_4)) + 2\tau(G_4) = 6 + 5\tau(G_4)$ . Finally, deletion of  $d$  from  $G_4$  leaves a graph with two spanning trees, while collapsing  $d$  produces a graph with three spanning trees. We conclude that  $\tau(G_4) = 5$ , so  $\tau(G) = 6 + 25 = 31$ , in agreement with the enumeration in 3.106.

Next we extend the preceding discussion to rooted spanning trees in digraphs.

**3.110. Definition: Rooted Spanning Trees.** Let  $G = (V, E, \epsilon)$  be a digraph, and let  $v_0 \in V$ . A *spanning tree of  $G$  rooted at  $v_0$*  is a rooted tree  $T$  with root  $v_0$  and vertex set  $V$  such that  $T$  (without the loop at  $v_0$ ) is a subgraph of  $G$ . Let  $\tau(G, v_0)$  be the number of spanning trees of  $G$  rooted at  $v_0$ .

The notions of edge deletion and contraction extend in a natural way to digraphs. This given, we have the following recursion for counting rooted spanning trees.

**3.111. Theorem: Rooted Spanning Tree Recursion.** Let  $v_0$  be a fixed vertex in a digraph  $G$ , and let  $z$  be a fixed edge leading into  $v_0$ . Let  $G_1$  be the digraph obtained from  $G$  by deleting  $z$ . Let  $G_2$  be the digraph obtained from  $G$  by collapsing  $z$ , and let the new “collapsed” vertex in  $G_2$  be  $v'_0$ . Then

$$\tau(G, v_0) = \tau(G_1, v_0) + \tau(G_2, v'_0).$$

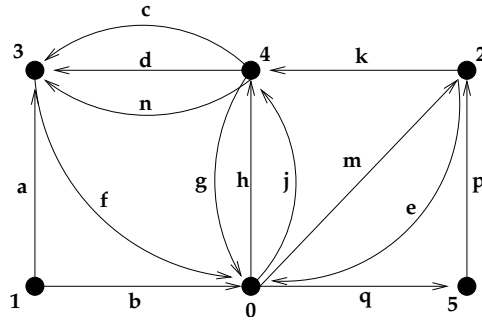
*Proof.* We modify the proof of 3.108. As before, the two terms on the right side count rooted spanning trees of  $G$  that do not contain  $z$  or do contain  $z$ . The reader should check that if  $T$  is a rooted spanning tree using the edge  $z$ , then the graph obtained from  $T$  by collapsing  $z$  is a rooted spanning tree of  $G_2$  rooted at  $v'_0$ . Similarly, adding  $z$  to the edge set of a rooted spanning tree of  $G_2$  rooted at  $v'_0$  produces a rooted spanning tree of  $G$  rooted at  $v_0$ .  $\square$

**3.112. Remark.** Our results for counting (undirected) spanning trees are special cases of the corresponding results for rooted spanning trees. For, given a graph  $G = (V, E, \epsilon)$ , consider the associated digraph obtained by replacing each  $e \in E$  by two directed edges going in opposite directions. Arguing as in 3.74, we see that there is a bijection between the set of rooted spanning trees of this digraph rooted at any given vertex  $v_0 \in V$  and the set of spanning trees of  $G$ . In the sequel, we shall only treat the case of digraphs.

---

## 3.17 Matrix-Tree Theorem

There is a remarkable determinant formula for the number of rooted spanning trees of a digraph. The formula uses the following modified version of the adjacency matrix of the digraph.


**FIGURE 3.25**

Digraph used to illustrate the matrix-tree theorem.

**3.113. Definition: Laplacian Matrix of a Digraph.** Let  $G$  be a loopless digraph on the vertex set  $V = \{v_0, v_1, \dots, v_n\}$ . The *Laplacian matrix* of  $G$  is the matrix  $L = (L_{ij} : 0 \leq i, j \leq n)$  such that  $L_{ii} = \text{outdeg}(v_i)$  and  $L_{ij}$  is the negative of the number of edges from  $v_i$  to  $v_j$  in  $G$ . We let  $L_0$  be the  $n \times n$  matrix obtained by erasing the row and column of  $L$  corresponding to  $v_0$ . The matrix  $L_0 = L_0(G)$  is called the *truncated Laplacian matrix* of  $G$  (relative to  $v_0$ ).

**3.114. Matrix-Tree Theorem.** With the notation of the preceding definition, we have

$$\tau(G, v_0) = \det(L_0(G)).$$

We prove the theorem after considering two examples.

**3.115. Example.** Let  $G$  be the digraph associated to the undirected graph in Figure 3.22. In this case,  $L_{ii}$  is the degree of vertex  $i$  in the undirected graph, and  $L_{ij}$  is minus the number of undirected edges between  $i$  and  $j$ . So

$$L = \begin{bmatrix} 4 & 0 & -2 & -1 & -1 \\ 0 & 3 & 0 & -2 & -1 \\ -2 & 0 & 3 & 0 & -1 \\ -1 & -2 & 0 & 3 & 0 \\ -1 & -1 & -1 & 0 & 3 \end{bmatrix}.$$

Striking out the row and column corresponding to vertex 0 leaves

$$L_0 = \begin{bmatrix} 3 & 0 & -2 & -1 \\ 0 & 3 & 0 & -1 \\ -2 & 0 & 3 & 0 \\ -1 & -1 & 0 & 3 \end{bmatrix}.$$

We compute  $\det(L_0) = 31$ , which agrees with our earlier calculation of  $\tau(G)$ .

**3.116. Example.** Consider the digraph  $G$  shown in Figure 3.25. We compute

$$L = \begin{bmatrix} 4 & 0 & -1 & 0 & -2 & -1 \\ -1 & 2 & 0 & -1 & 0 & 0 \\ -1 & 0 & 2 & 0 & -1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & -3 & 4 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}, \quad L_0 = \begin{bmatrix} 2 & 0 & -1 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -3 & 4 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix},$$

and  $\det(L_0) = 16$ . So  $G$  has 16 spanning trees rooted at 0, as one may confirm by direct enumeration. We will use the matrix  $L_0$  as a running example in the proof below.

**3.117. Proof of the Matrix-Tree Theorem.** Write  $L_0 = L_0(G)$ . First we prove that  $\tau(G, v_0) = \det(L_0)$  in the case where  $\text{indeg}(v_0) = 0$ . If  $v_0$  is the only vertex of  $G$ , then  $\tau(G, v_0) = 1$  and  $\det(L_0) = 1$  by the convention that the determinant of a  $0 \times 0$  matrix is 1. Otherwise,  $\tau(G, v_0)$  is zero, and  $L_0$  is a nonempty matrix. Using the condition  $\text{indeg}(v_0) = 0$  and the definition of  $L_0$ , one sees that every row of  $L_0$  sums to zero. Therefore, letting  $\vec{u}$  be a column vector of  $n$  ones, we have  $L_0\vec{u} = \vec{0}$ , so that  $L_0$  is singular and  $\det(L_0) = 0$ .

For the general case, we argue by induction on the number of edges in  $G$ . The case where  $G$  has no edges is covered by the previous paragraph. The only case left to consider occurs when  $\text{indeg}(v_0) > 0$ . Let  $e$  be a fixed edge in  $G$  that leads from some  $v_i$  to  $v_0$ . Let  $G_1$  be the graph obtained from  $G$  by deleting  $e$ , and let  $G_2$  be the graph obtained from  $G$  by collapsing  $e$ . Both graphs have fewer edges than  $G$ , so the induction hypothesis tells us that

$$\tau(G_1, v_0) = \det(L_0(G_1)) \text{ and } \tau(G_2, v'_0) = \det(L_0(G_2)), \quad (3.4)$$

where  $v'_0$  is the new vertex created after collapsing  $e$ . Using 3.111, we conclude that

$$\tau(G, v_0) = \det(L_0(G_1)) + \det(L_0(G_2)). \quad (3.5)$$

Next, let us evaluate the determinant  $\det(L_0(G))$ . We will use the fact that the determinant of a matrix is a *linear* function of each row of the matrix. More precisely, for a fixed matrix  $A$  and row index  $i$ , let  $A[y]$  denote the matrix  $A$  with the  $i$ th row replaced by the row vector  $y$ ; then  $\det(A[y + z]) = \det(A[y]) + \det(A[z])$  for all  $y, z$ . This linearity property can be proved directly from the definition of the determinant (see 9.37 and 9.45 below). To apply this result, write the  $i$ th row of  $L_0 = L_0(G)$  in the form  $y + z$ , where  $z = (0, 0, \dots, 1, 0, \dots, 0)$  has a one in position  $i$ . Then

$$\det(L_0(G)) = \det(L_0[y]) + \det(L_0[z]). \quad (3.6)$$

For example, if  $G$  is the digraph in Figure 3.25 and  $e$  is the edge from 2 to 0 (so  $i = 2$ ), then  $y = (0, 1, 0, -1, 0)$ ,  $z = (0, 1, 0, 0, 0)$ ,

$$L_0[y] = \begin{bmatrix} 2 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -3 & 4 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix}, \quad L_0[z] = \begin{bmatrix} 2 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -3 & 4 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix}.$$

Comparing equations (3.5) and (3.6), we see that it suffices to prove  $\det(L_0(G_1)) = \det(L_0[y])$  and  $\det(L_0(G_2)) = \det(L_0[z])$ .

How does the removal of  $e$  from  $G$  affect  $L(G)$ ? Answer: The  $i, i$ -entry drops by 1, while the  $i, 0$ -entry increases by 1. Since the zeroth column is ignored in the truncated Laplacian, we see that we can obtain  $L_0(G_1)$  from  $L_0(G)$  by decrementing the  $i, i$ -entry by 1. In other words,  $L_0(G_1) = L_0[y]$ , and hence  $\det(L_0(G_1)) = \det(L_0[y])$ .

Next, let us calculate  $\det(L_0[z])$  by expanding the determinant along row  $i$ . The only nonzero entry in this row is the 1 in the diagonal position, so  $\det(L_0[z]) = (-1)^{i+i} \det(M) = \det(M)$ , where  $M$  is the matrix obtained from  $L_0[z]$  (or equivalently, from  $L_0$ ) by erasing row  $i$  and column  $i$ . In our running example,

$$M = \begin{bmatrix} 2 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -3 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We claim that  $M = L_0(G_2)$ , which will complete the proof. Consider the  $k, j$ -entry of  $M$ , where  $k, j \in \{0, 1, \dots, n\} \sim \{0, i\}$ . If  $k = j$ , this entry is  $\text{outdeg}_G(v_j)$ , which equals  $\text{outdeg}_{G_2}(v_j)$  because  $v_j$  is not  $v_0$ ,  $v_i$ , or  $v'_0$ . For the same reason, if  $k \neq j$ , the  $k, j$ -entry of  $M$  is minus the number of edges from  $v_k$  to  $v_j$ , which is the same in  $G$  and  $G_2$ .

### 3.18 Eulerian Tours

**3.118. Definition: Eulerian Tours.** Let  $G = (V, E, \epsilon)$  be a digraph. An *Eulerian tour* in  $G$  is a walk  $W = (v_0, e_1, v_1, e_2, v_2, \dots, e_n, v_n)$  such that  $W$  visits every vertex in  $V$ , and  $W$  uses every edge in  $E$  exactly once. Such a tour is called *closed* iff  $v_n = v_0$ .

**3.119. Example.** Consider the digraph  $G$  shown in Figure 3.26. Here is one closed Eulerian tour of  $G$ :

$$W_1 = (0, m, 2, l, 5, e, 1, a, 3, c, 4, b, 3, d, 5, f, 4, g, 5, k, 0, i, 4, h, 5, j, 0).$$

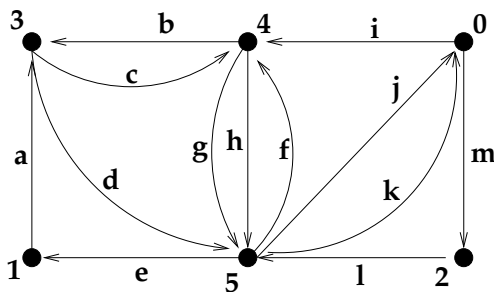
To specify the tour, it suffices to list only the edges in the tour. For instance, here is the edge sequence of another closed Eulerian tour of  $G$ :

$$W_2 = (i, g, e, a, d, f, b, c, h, j, m, l, k).$$

**3.120. Example.** Consider the digraph  $G$  shown in Figure 3.2. This graph does not have any closed Eulerian tours, since there is no way to reach vertex 6 from the other vertices. Even if we delete vertex 6 from the graph, there will still be no closed Eulerian tours. For, there is no way that a tour can use both edges leaving vertex 2, since only one edge enters vertex 2.

The previous example indicates two necessary conditions for a digraph to have a closed Eulerian tour: the digraph must be connected, and also *balanced* in the sense that  $\text{indeg}(v) = \text{outdeg}(v)$  for every vertex  $v$ . We now show that these necessary conditions are also sufficient to guarantee the existence of a closed Eulerian tour.

**3.121. Theorem: Existence of Closed Eulerian Tours.** A digraph  $G = (V, E, \epsilon)$  has a closed Eulerian tour iff  $G$  is connected and balanced.



**FIGURE 3.26**

Digraph used to illustrate Eulerian tours.



*Proof.* First suppose  $G$  has a closed Eulerian tour  $W$  starting at  $v_0$ . Since  $W$  visits every vertex, we can obtain a walk from any vertex to any other vertex by following suitable edges of  $W$ . So  $G$  is connected. Next, let  $v$  be any vertex of  $G$ . The walk  $W$  arrives at  $v$  via an incoming edge exactly as often as the walk leaves  $v$  via an outgoing edge; this is true even if  $v = v_0$ . Since the walk uses every edge exactly once, it follows that  $\text{indeg}(v) = \text{outdeg}(v)$ .

Conversely, assume that  $G$  is connected and balanced. Let  $W = (v_0, e_1, v_1, \dots, e_n, v_n)$  be a walk of maximum length in  $G$  that never repeats an edge. We claim that  $v_n = v_0$ . For, if not,  $W$  enters vertex  $v_n$  one more time than it leaves  $v_n$ . Since  $\text{indeg}(v_n) = \text{outdeg}(v_n)$ , there must be an outgoing edge from  $v_n$  that has not been used by  $W$ . So we could use this edge to extend  $W$ , contradicting maximality. Next, we claim that  $W$  uses *every* edge of  $G$ . If not, let  $e$  be an edge not used by  $W$ . Since  $G$  is connected, we can find such an edge that is incident to one of the vertices  $v_i$  visited by  $W$ . Since  $v_n = v_0$ , we can cyclically shift the walk  $W$  to get a new walk  $W' = (v_i, e_{i+1}, v_{i+1}, \dots, e_n, v_n = v_0, e_1, \dots, e_i, v_i)$  that starts and ends at  $v_i$ . By adding the edge  $e$  to the beginning or end of this walk (depending on its direction), we could again produce a longer walk than  $W$  with no repeated edges, violating maximality. Finally,  $W$  must visit every vertex of  $G$ , since  $W$  uses every edge of  $G$  and (unless  $G$  has one vertex and no edges) every vertex has an edge leaving it.  $\square$

Our goal in the rest of this section is to prove the following formula for the number of closed Eulerian tours in  $G$  starting at a given vertex  $v_0$ . Recall that  $\tau(G, v_0)$  is the number of rooted spanning trees of  $G$  rooted at  $v_0$ .

**3.122. Theorem: Counting Eulerian Tours.** Let  $G = (V, E, \epsilon)$  be a connected, balanced digraph. For each  $v_0 \in V$ , the number of closed Eulerian tours of  $G$  starting at  $v_0$  is

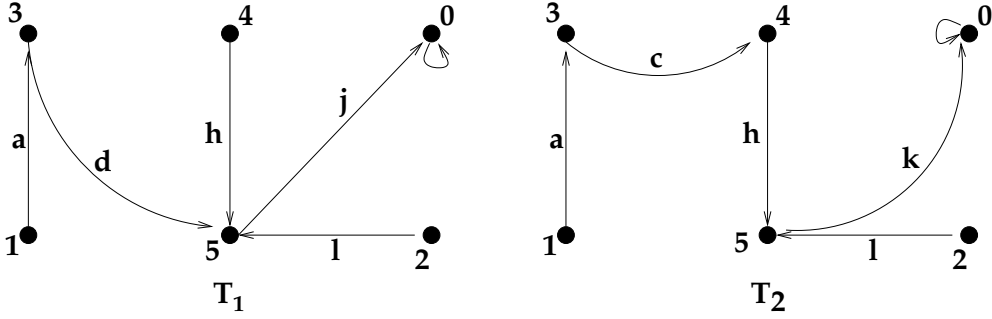
$$\tau(G, v_0) \cdot \text{outdeg}(v_0)! \cdot \prod_{v \neq v_0} (\text{outdeg}(v) - 1)! \quad (3.7)$$

Let  $\{v_0, v_1, \dots, v_n\}$  be the vertex set of  $G$ . Let  $X$  be the set of all closed Eulerian tours of  $G$  starting at  $v_0$ . Let  $\text{SpTr}(G, v_0)$  be the set of spanning trees of  $G$  rooted at  $v_0$ . Let  $Y$  be the set of all tuples  $(T, w_0, w_1, w_2, \dots, w_n)$  where:  $T \in \text{SpTr}(G, v_0)$ ;  $w_0$  is a permutation of all the edges leaving  $v_0$ ; and, for  $1 \leq i \leq n$ ,  $w_i$  is a permutation of those edges leaving  $v_i$  other than the unique outgoing edge from  $v_i$  that belongs to  $T$  (see 3.42). By the product rule, the cardinality of  $Y$  is given by the right side of (3.7). So it will suffice to define a bijection  $f : X \rightarrow Y$ .

Given an Eulerian tour  $W \in X$ , define  $f(W) = (T, w_0, \dots, w_n)$  as follows. For each  $i$  between 0 and  $n$ , let  $w'_i$  be the permutation of *all* edges leading out of  $v_i$ , taken in the order in which they occur in the walk  $W$ . Call  $w'_i$  the *departure word* of vertex  $v_i$ . Next, set  $w_0 = w'_0$  and for  $i > 0$ , let  $w_i$  be the word  $w'_i$  with the last symbol erased. Finally, let  $T$  be the subgraph of  $G$  whose edges are given by the last symbols of  $w'_1, \dots, w'_n$ , augmented by a loop edge at  $v_0$ . It is not immediately evident that  $T \in \text{SpTr}(G, v_0)$ ; we will prove this shortly.

Next we define a map  $g : Y \rightarrow X$  that will be the two-sided inverse of  $f$ . Fix  $(T, w_0, \dots, w_n) \in Y$ . For every  $i > 0$ , form  $w'_i$  by appending the unique edge of  $T$  leaving  $v_i$  to the end of the word  $w_i$ ; let  $w'_0 = w_0$ . Starting at  $v_0$ , we use the words  $w'_i$  to build a walk through  $G$ , one edge at a time, as follows. If we are currently at some vertex  $v_i$ , use the next unread symbol in  $w'_i$  to determine which edge to follow out of  $v_i$ . Repeat this process until the walk reaches a vertex in which all the outgoing edges have already been used. The resulting walk  $W$  is  $g(T, w_0, \dots, w_n)$ . The edges occurring in  $W$  are pairwise distinct, but it is not immediately evident that  $W$  must use *all* edges of  $G$ ; we will prove this shortly.

Once we check that  $f$  and  $g$  map into their stated codomains, the definitions just given show that  $f \circ g$  and  $g \circ f$  are both identity maps. Before proving that  $f$  maps into  $Y$  and  $g$  maps into  $X$ , let us consider an example.

**FIGURE 3.27**

Rooted spanning trees associated to Eulerian tours.

**3.123. Example.** We continue the analysis of Eulerian tours in the digraph  $G$  from 3.119. The walk  $W_1$  in that example has departure words  $w'_0 = mi$ ,  $w'_1 = a$ ,  $w'_2 = l$ ,  $w'_3 = cd$ ,  $w'_4 = bgh$ , and  $w'_5 = efkj$ . Therefore,

$$f(W_1) = (T_1, mi, \cdot, \cdot, c, bg, efk),$$

where  $\cdot$  denotes an empty word and  $T_1$  is the graph shown on the left in Figure 3.27. Similarly, for  $W_2$  we compute  $w'_0 = im$ ,  $w'_1 = a$ ,  $w'_2 = l$ ,  $w'_3 = dc$ ,  $w'_4 = gbh$ ,  $w'_5 = efjk$ , and

$$f(W_2) = (T_2, im, \cdot, \cdot, d, gb, efj).$$

Let us now calculate  $g((T_1, im, \cdot, \cdot, c, bg, fke))$ . First, we use the edges of  $T_1$  to recreate the departure words  $w'_0 = im$ ,  $w'_1 = a$ ,  $w'_2 = l$ ,  $w'_3 = cd$ ,  $w'_4 = bgh$ , and  $w'_5 = fkej$ . We then use these words to guide our tour through the graph. We begin with  $0, i, 4$ , since  $i$  is the first letter of  $w'_0$ . Consulting  $w'_4$  next, we follow edge  $b$  to vertex 3, then edge  $c$  to vertex 4, then edge  $g$  to vertex 5, and so on. We obtain the tour

$$W_3 = (0, i, 4, b, 3, c, 4, g, 5, f, 4, h, 5, k, 0, m, 2, l, 5, e, 1, a, 3, d, 5, j, 0).$$

Similarly, the reader may check that

$$g((T_2, mi, \cdot, \cdot, d, bg, jfe)) = (m, l, j, i, b, d, f, g, e, a, c, h, k).$$

To complete the proof of 3.122, we must prove two things. First, to show that  $f(W) \in Y$  for all  $W \in X$ , we must show that the digraph  $T$  obtained from the last letters of the departure words  $w'_i$  ( $i > 0$ ) is a rooted spanning tree of  $G$  rooted at  $v_0$ . Since  $W$  visits every vertex of  $G$ , the definition of  $T$  shows that  $\text{outdeg}_T(v_i) = 1$  for all  $i \geq 0$ . We need only show that  $T$  has no cycles other than the loop at  $v_0$  (see 3.42). We can view the tour  $W$  as a certain permutation of all the edges in  $G$ . Let us show that if  $e, h$  are two non-loop edges in  $T$  with  $\epsilon(e) = (x, y)$  and  $\epsilon(h) = (y, z)$ , then  $e$  must precede  $h$  in the permutation  $W$ . Note that  $y$  cannot be  $v_0$ , since the only outgoing edge from  $v_0$  in  $T$  is a loop edge. Thus, when the tour  $W$  uses the edge  $e$  to enter  $y$ , the following edge in the tour exists and is an outgoing edge from  $y$ . Since  $h$  is, by definition, the *last* such edge used by the tour,  $e$  must precede  $h$  in the tour. Now suppose  $(z_0, e_1, z_1, \dots, e_n, z_n)$  is a cycle in  $T$  that is not the 1-cycle at  $v_0$ . Using the previous remark repeatedly, we see that  $e_i$  precedes  $e_{i+1}$  in  $W$  for all  $i$ , and also  $e_n$  precedes  $e_1$  in  $W$ . These statements imply that  $e_1$  precedes itself in  $W$ , which is absurd. We conclude that  $f(W) \in Y$ .

Second, we must show that  $g$  maps  $Y$  into  $X$ . Fix  $(T, w_0, \dots, w_n) \in Y$  and  $W = g(T, w_0, \dots, w_n)$ , and let  $w'_i$  be the departure words constructed from  $T$  and the  $w_i$ 's. We know from the definition of  $g$  that  $W$  is a walk in  $G$  starting at  $v_0$  that never repeats an edge. We must show that  $W$  ends at  $v_0$  and uses every edge in  $G$ . Suppose, at some stage in the construction of  $W$ , that  $W$  has just reached  $v_i$  for some  $i > 0$ . Then  $W$  has entered  $v_i$  one more time than it has left  $v_i$ . Since  $G$  is balanced, there must exist an unused outgoing edge from  $v_i$ . This edge corresponds to an unused letter in  $w'_i$ . So  $W$  does not end at  $v_i$ . The only possibility is that  $W$  ends at the starting vertex  $v_0$ .

To prove that  $W$  uses every edge of  $G$ , we claim that it is enough to prove that  $W$  uses every non-loop edge of  $T$ . For, consider a vertex  $v \neq v_0$  of  $G$ . If  $W$  uses the unique outgoing edge from  $v$  that is part of  $T$ , then  $W$  must have previously used all other outgoing edges from  $v$ , by definition of  $W$ . Since  $W$  ends at  $v_0$ ,  $W$  certainly uses all outgoing edges from  $v_0$ . All edges are accounted for in this way, proving the claim.

Finally, to get a contradiction, assume that some edge  $e$  in  $T$  from  $x$  to  $y$  is not used by  $W$ . Since  $T$  is a rooted tree rooted at  $v_0$ , we can choose such an  $e$  so that the distance from  $y$  to  $v_0$  through edges in  $T$  is minimal. If  $y \neq v_0$ , minimality implies that the unique edge leading out of  $y$  in  $T$  does belong to  $W$ . Then, as noted in the last paragraph, every outgoing edge from  $y$  in  $G$  is used in  $W$ . Since  $G$  is balanced, every incoming edge into  $y$  in  $G$  must also appear in  $W$ , contradicting the assumption that  $e$  is not used by  $W$ . On the other hand, if  $y = v_0$ , we see similarly that  $W$  uses every outgoing edge from  $y$  in  $G$  and hence every incoming edge to  $y$  in  $G$ . Again, this contradicts the assumption that  $e$  is not in  $W$ . This completes the proof of 3.122.

## Summary

Table 3.1 contains brief definitions of the terminology from graph theory used in this chapter.

- *Facts about Matrix Multiplication.* If  $A_1, \dots, A_s$  are matrices such that  $A_t$  is  $n_{t-1} \times n_t$ , then the  $i, j$ -entry of the product  $A_1 A_2 \cdots A_s$  is

$$\sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \cdots \sum_{k_{s-1}=1}^{n_{s-1}} A_1(i, k_1) A_2(k_1, k_2) A_3(k_2, k_3) \cdots A_s(k_{s-1}, j).$$

If  $A^s = 0$  (i.e.,  $A$  is nilpotent), then  $I - A$  is invertible, and

$$(I - A)^{-1} = I + A + A^2 + A^3 + \cdots + A^{s-1}.$$

This formula applies (with  $s = n$ ) when  $A$  is a strictly upper or lower triangular  $n \times n$  matrix.

- *Adjacency Matrices and Walks.* Given a graph or digraph  $G$  with vertex set  $\{v_1, \dots, v_n\}$ , the adjacency matrix of  $G$  is the matrix  $A$  such that  $A(i, j)$  is the number of edges from  $v_i$  to  $v_j$  in  $G$ . For all  $s \geq 0$ ,  $A^s(i, j)$  is the number of walks in  $G$  of length  $s$  from  $v_i$  to  $v_j$ .  $G$  is a DAG iff  $A^n = 0$ , in which case  $A$  will be strictly lower-triangular under a suitable ordering of the vertices. When  $G$  is a DAG,  $(I - A)^{-1}(i, j)$  is the total number of paths (or walks) from  $v_i$  to  $v_j$ .
- *Degree-Sum Formulas.* For a digraph  $G = (V, E, \epsilon)$ ,

$$\sum_{v \in V} \text{indeg}_G(v) = |E| = \sum_{v \in V} \text{outdeg}_G(v).$$

**TABLE 3.1**

Terminology used in graph theory.

Term	Brief Definition
graph	$(V, E, \epsilon)$ where $\epsilon(e) = \{v, w\}$ means edge $e$ has endpoints $v, w$
digraph	$(V, E, \epsilon)$ where $\epsilon(e) = (v, w)$ means edge $e$ goes from $v$ to $w$
simple graph	graph with no loops or multiple edges
simple digraph	digraph with no multiple edges
$G \cong H$	$G$ becomes $H$ under suitable renaming of vertices and edges
walk	$(v_0, e_1, v_1, \dots, e_s, v_s)$ where each $e_i$ is an edge from $v_{i-1}$ to $v_i$
closed walk	walk starts and ends at same vertex
path	walk visiting distinct vertices
cycle	closed walk visiting distinct vertices and edges, except at end
DAG	digraph with no cycles
$\text{indeg}_G(v)$	number of edges leading to $v$ in digraph $G$
$\text{outdeg}_G(v)$	number of edges leading from $v$ in digraph $G$
$\text{deg}_G(v)$	number of edges incident to $v$ in graph $G$ (loops count as 2)
isolated vertex	vertex of degree zero
leaf	vertex of degree one
functional digraph	simple digraph with $\text{outdeg}(v) = 1$ for all vertices $v$
cyclic vertex	vertex in functional digraph that belongs to a cycle
rooted tree	functional digraph with a unique cyclic vertex (the root)
$G$ is connected	for all $u, v \in V(G)$ , there is a walk in $G$ from $u$ to $v$
cut-edge of $G$	edge belonging to no cycle of the graph $G$
forest	graph with no cycles
acyclic graph	graph with no cycles
tree	connected graph with no cycles
proper coloring	map $f : V(G) \rightarrow C$ assigning unequal colors to adjacent vertices
$\chi_G(x)$	number of proper colorings of $G$ using $x$ available colors
chromatic number	least $x$ with $\chi_G(x) > 0$
subgraph of $G$	graph $G'$ with $V(G') \subseteq V(G)$ , $E(G') \subseteq E(G)$ (same endpoints)
induced subgraph	subgraph $G'$ where all edges in $G$ with ends in $V(G')$ are kept
spanning tree of $G$	subgraph of $G$ that is a tree using all vertices
$\tau(G)$	number of spanning trees of $G$
rooted spanning tree	rooted tree using all vertices of a digraph
$\tau(G, v_0)$	number of rooted spanning trees of $G$ with root $v_0$
Eulerian tour	walk visiting each vertex that uses every edge once

For a graph  $G = (V, E, \epsilon)$ ,

$$\sum_{v \in V} \deg_G(v) = 2|E|.$$

- *Functional Digraphs.* For a finite set  $X$ , every function  $f : X \rightarrow X$  has an associated functional digraph with vertex set  $X$  and edge set  $\{(x, f(x)) : x \in X\}$ . Every functional digraph decomposes uniquely into one or more disjoint cycles together with disjoint rooted trees rooted at the vertices on these cycles. For each vertex  $x_0$  in a functional digraph, there exist unique walks of each length  $k$  starting at  $x_0$ , which are found by repeatedly following the unique outgoing edge from the current vertex. Such walks eventually reach a cycle in the functional digraph.
- *Cycle Structure of Permutations.* For  $X$  finite, a map  $f : X \rightarrow X$  is a bijection iff the functional digraph of  $f$  is a disjoint union of directed cycles. The signless Stirling number of the first kind,  $s'(n, k)$ , counts the number of bijections  $f$  on an  $n$ -element set such that the functional digraph of  $f$  has  $k$  cycles. We have

$$s'(n, k) = s'(n-1, k-1) + (n-1)s'(n-1, k) \quad (0 < k < n).$$

- *Connectedness and Components.* The vertex set of any graph or digraph  $G$  is the disjoint union of connected components. Two vertices belong to the same component iff each vertex is reachable from the other by a walk.  $G$  is connected iff there is only one component iff for all  $u, v \in V(G)$  there exists at least one *path* from  $u$  to  $v$  in  $G$ . Deleting a cut-edge splits a component of  $G$  in two, whereas deleting a non-cut-edge has no effect on components.
- *Forests.* A graph  $G$  is a forest (acyclic) iff  $G$  has no loops and for each  $u, v \in V(G)$ , there is at most one path from  $u$  to  $v$ . A forest with  $n$  vertices and  $k$  edges has  $n - k$  components.
- *Trees.* The following conditions on an  $n$ -vertex simple graph  $G$  are equivalent and characterize trees: (a)  $G$  is connected with no cycles; (b)  $G$  is connected with at most  $n - 1$  edges; (c)  $G$  is acyclic with at least  $n - 1$  edges; (d) for all  $u, v \in V(G)$ , there exists a unique path in  $G$  from  $u$  to  $v$ . An  $n$ -vertex tree has  $n - 1$  edges and (for  $n > 1$ ) at least two leaves. Pruning any leaf from a tree gives another tree with one less vertex and one less edge.
- *Tree Enumeration Results.* There are  $n^{n-2}$  trees with vertex set  $\{1, 2, \dots, n\}$ . There are  $n^{n-2}$  rooted trees on this vertex set rooted at 1. For  $d_1 + \dots + d_n = 2(n-1)$ , there are  $\binom{n-2}{d_1-1, \dots, d_n-1}$  trees on this vertex set with  $\deg(j) = d_j$  for all  $j$ . Bijective proofs of these facts use the following ideas:
  - Functions on  $\{1, 2, \dots, n\}$  fixing 1 and  $n$  correspond to rooted trees by arranging the cycles of the functional digraph in a certain order, breaking “back edges,” and linking the cycles to get a tree (see Figures 3.7 and 3.8).
  - Trees correspond to rooted trees by directing each edge of the tree towards the desired root vertex.
  - Trees with  $\deg(j) = d_j$  correspond to words in  $\mathcal{R}(1^{d_1-1} \dots n^{d_n-1})$  by repeatedly pruning the largest leaf and appending the leaf’s neighbor to the end of the word.

- *Terms and Ordered Trees.* For every term  $T$ , there exist a unique integer  $n \geq 0$  and unique terms  $T_1, \dots, T_n$  such that  $T = nT_1T_2 \cdots T_n$ . A word  $w_1 \cdots w_s$  is a term iff  $w_1 + \cdots + w_i - i \geq 0$  for all  $i < s$  and  $w_1 + \cdots + w_s - s = -1$ . No proper prefix of a term is a term. Terms correspond bijectively to ordered trees.
- *Lists of Terms and Ordered Forests.* Every list of terms has the form  $T_1 \cdots T_n$  for some unique integer  $n \geq 0$  and unique terms  $T_1, \dots, T_n$ . A word  $w_1 \cdots w_s$  is a list of  $n$  terms iff  $w_1 + \cdots + w_i - i > -n$  for all  $i < s$  and  $w_1 + \cdots + w_s - s = -n$ . Lists of terms correspond bijectively to ordered forests.
- *Cycle Lemma and Enumeration of Lists of Terms.* For a word  $w = w_1 \cdots w_s$  with  $w_1 + \cdots + w_s - s = -n$ , there exist exactly  $n$  cyclic shifts of  $w$  that are lists of  $n$  terms. Consequently, the number of lists of  $n$  terms using  $k_i$  copies of  $i$  (for  $0 \leq i \leq t$ ) is

$$\frac{n}{s} \binom{s}{k_0, k_1, \dots, k_s},$$

where  $s = \sum_{i=0}^t k_i$  and  $k_0 = n + \sum_{i=1}^t (i-1)k_i$ .

- *Chromatic Polynomials.* For any edge  $e$  in a simple graph  $G$ , the chromatic function of  $G$  satisfies the recursion  $\chi_G = \chi_{G \sim \{e\}} - \chi_{G_e}$ , where  $G \sim \{e\}$  is  $G$  with  $e$  deleted, and  $G_e$  is  $G$  with  $e$  collapsed. It follows that  $\chi_G(x)$  is a polynomial function of  $x$ . The signed Stirling numbers of the first kind,  $s(n, k)$ , are the coefficients in the chromatic polynomial for an  $n$ -vertex graph with an edge between each pair of vertices.
- *Spanning Tree Recursion.* For any edge  $e$  in a graph  $G$ , the number  $\tau(G)$  of spanning trees of  $G$  satisfies the recursion  $\tau(G) = \tau(G \sim \{e\}) + \tau(G_e)$ , where  $G \sim \{e\}$  is  $G$  with  $e$  deleted, and  $G_e$  is  $G$  with  $e$  collapsed. A similar recursion holds for rooted spanning trees of a digraph.
- *Matrix-Tree Theorem.* Given a digraph  $G$  and  $v_0 \in V(G)$ , let  $L_{ii} = \text{outdeg}_G(v_i)$ , let  $-L_{ij}$  be the number of edges from  $i$  to  $j$  in  $G$ , and let  $L_0$  be the matrix obtained from  $(L_{ij})$  by erasing the row and column indexed by  $v_0$ . Then  $\det(L_0)$  is the number  $\tau(G, v_0)$  of rooted spanning trees of  $G$  with root  $v_0$ .
- *Eulerian Tours.* A digraph  $G$  has a closed Eulerian tour iff  $G$  is connected and balanced (indegree equals outdegree at every vertex). In this case, the number of such tours starting at  $v_0$  is

$$\tau(G, v_0) \cdot \text{outdeg}_G(v_0)! \cdot \prod_{v \neq v_0} (\text{outdeg}_G(v) - 1)!.$$

The proof associates to each tour a rooted spanning tree built from the last departure edge from each vertex, together with (truncated) departure words for each vertex giving the order in which the tour used the other outgoing edges.

## Exercises

**3.124.** Draw pictures of the following simple graphs, which have the indicated nicknames.

- the *claw*  $C = (\{1, 2, 3, 4\}, \{\{1, 2\}, \{1, 3\}, \{1, 4\}\})$ ;
- the *paw*  $P = (\{1, 2, 3, 4\}, \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}\})$ ;

- (c) the *kite*  $K = (\{1, 2, 3, 4\}, \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\})$ ;  
 (d) the *bull*  $B = (\{1, 2, 3, 4, 5\}, \{\{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 4\}, \{2, 5\}\})$ .  
 (e) the  $n$ -*path*  $P_n = (\{1, 2, \dots, n\}, \{\{i, i+1\} : 1 \leq i < n\})$ .  
 (f) the  $n$ -*cycle*  $C_n = (\{1, 2, \dots, n\}, \{\{i, i+1\} : 1 \leq i < n\} \cup \{\{1, n\}\})$ , where  $n \geq 3$ .  
 (g) the *complete graph*  $K_n = (\{1, 2, \dots, n\}, \{\{i, j\} : 1 \leq i < j \leq n\})$ .

**3.125.** Let  $V$  be an  $n$ -element set. (a) How many simple graphs have vertex set  $V$ ? (b) How many simple digraphs have vertex set  $V$ ?

**3.126.** Let  $V$  and  $E$  be sets with  $|V| = n$  and  $|E| = m$ . (a) How many digraphs have vertex set  $V$  and edge set  $E$ ? (b) How many graphs have vertex set  $V$  and edge set  $E$ ?

**3.127.** Let  $V$  be an  $n$ -element set. Define a bijection between the set of simple graphs with vertex set  $V$  and the set of symmetric, irreflexive binary relations on  $V$ . Conclude that simple graphs can be viewed as certain kinds of simple digraphs.

**3.128.** Let  $G$ ,  $H$ , and  $K$  be graphs (resp. digraphs). (a) Prove  $G \cong G$ . (b) Prove  $G \cong H$  implies  $H \cong G$ . (c) Prove  $G \cong H$  and  $H \cong K$  imply  $G \cong K$ . Thus, graph isomorphism is an equivalence relation on any given set of graphs (resp. digraphs).

**3.129.** Find all isomorphism classes of simple graphs with at most four vertices.

**3.130.** Find the adjacency matrices for the graphs in 3.124.

**3.131.** Let  $G$  be the simple graph in Figure 3.10. For  $1 \leq k \leq 8$ , find the number of walks in  $G$  from vertex 1 to vertex 10.

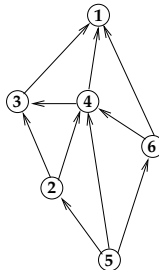
**3.132.** Let  $G$  be the graph in Figure 3.22. Find the number of walks in  $G$  of length 5 between each pair of vertices.

**3.133.** Let  $G$  be the digraph in Figure 3.25. Find the number of closed walks in  $G$  of length 10 that begin at vertex 0.

**3.134.** (a) Show that a graph  $G$  with a closed walk of odd length must have a cycle of odd length. (b) If  $G$  has a closed walk of even length, must  $G$  have a cycle?

**3.135.** Let  $G$  be a graph with adjacency matrix  $A$ . (a) Find a formula for the number of *paths* in  $G$  of length 2 from  $v_i$  to  $v_j$ . (b) Find a formula for the number of paths in  $G$  of length 3 from  $v_i$  to  $v_j$ .

**3.136.** Consider the DAG  $G$  shown here.



(a) Find all total orderings of the vertices for which the adjacency matrix of  $G$  is strictly lower-triangular. (b) How many paths in  $G$  go from vertex 5 to vertex 1?

**3.137.** An irreflexive, transitive binary relation on a set  $X$  is called a *strict partial order* on  $X$ . Given a strict partial order  $R$  on a finite set  $X$ , show that the simple digraph  $(X, R)$  is a DAG.

**3.138.** For each of the following sets  $X$  and strict partial orders  $R$ , draw the associated DAG (see 3.137) and calculate the number of paths from the smallest element to the largest element of the partially ordered set.

(a)  $X = \{1, 2, 3, 4, 5\}$  under the ordering  $1 < 2 < 3 < 4 < 5$ .

(b)  $X = \mathcal{P}(\{1, 2, 3\})$ , and  $(S, T) \in R$  iff  $S \subsetneq T$ .

(c)  $X$  is the set of positive divisors of 60, and  $(a, b) \in R$  iff  $a < b$  and  $a$  divides  $b$ .

**3.139.** Let  $X = \{1, 2, \dots, n\}$  ordered by  $1 < 2 < \dots < n$ . In the associated DAG (see 3.137), how many paths go from 1 to  $n$ ? Can you find a combinatorial (not algebraic) proof of your answer?

**3.140.** Let  $X$  be the set of subsets of  $\{1, 2, \dots, n\}$  ordered by (strict) set inclusion. In the associated DAG (see 3.137), how many paths go from  $\emptyset$  to  $\{1, 2, \dots, n\}$ ?

**3.141.** Given a digraph  $G$ , construct a simple digraph  $H$  as follows. The vertices of  $H$  are the strong components of  $G$ . Given  $C, D \in V(H)$  with  $C \neq D$ , there is an edge from  $C$  to  $D$  in  $H$  iff there exists  $c \in C$  and  $d \in D$  such that there is an edge from  $c$  to  $d$  in  $G$ . (a) Prove that  $H$  is a DAG. (b) Conclude that some strong component  $C$  of  $G$  has no incoming edges from outside  $C$ , and some strong component  $D$  has no outgoing edges. (c) Draw the DAGs associated to the digraph  $G_3$  in Figure 3.1 and the functional digraph in Figure 3.5.

**3.142.** (a) Find the degree sequence for the graph in Figure 3.10, and verify 3.34 in this case. (b) Compute the indegrees and outdegrees at each vertex of the digraph in Figure 3.25, and verify 3.31 in this case.

**3.143.** Find necessary and sufficient conditions for a multiset  $[d_1, d_2, \dots, d_n]$  to be the degree sequence of a graph  $G$ .

**3.144.** Consider the cycle graph  $C_n$  defined in 3.124. (a) What is  $\deg(C_n)$ ? (b) Show that any connected graph with the degree sequence in (a) must be isomorphic to  $C_n$ . (c) How many graphs with vertex set  $\{1, 2, \dots, n\}$  are isomorphic to  $C_n$ ? (d) How many isomorphism classes of graphs have the same degree sequence as  $C_n$ ? (e) How many isomorphism classes of *simple* graphs have the same degree sequence as  $C_n$ ?

**3.145.** Consider the path graph  $P_n$  defined in 3.124. (a) What is  $\deg(P_n)$ ? (b) Show that any connected graph with the degree sequence in (a) must be isomorphic to  $P_n$ . (c) How many graphs with vertex set  $\{1, 2, \dots, n\}$  are isomorphic to  $P_n$ ? (d) How many isomorphism classes of graphs have the same degree sequence as  $P_n$ ?

**3.146.** Find two simple graphs  $G$  and  $H$  with the smallest possible number of vertices, such that  $\deg(G) = \deg(H)$  but  $G \not\cong H$ .

**3.147.** Prove or disprove: there exists a simple graph  $G$  with more than one vertex such that the degree sequence  $\deg(G)$  contains no repetitions.

**3.148.** Prove or disprove: there exists a graph  $G$  with no loops and more than one vertex such that the degree sequence  $\deg(G)$  contains no repetitions.

**3.149.** Given a graph  $G = (V, E, \epsilon)$ , we can encode the endpoint function  $\epsilon$  by a  $|V| \times |E|$  matrix  $M$ , with rows indexed by  $V$  and columns indexed by  $E$ , such that  $M(v, e)$  is 2 if  $e$  is a loop edge at  $v$ , 1 if  $e$  is a non-loop edge incident to  $v$ , and 0 otherwise.  $M$  is called the *incidence matrix* of  $G$ . Prove the degree-sum formula 3.34 by computing the sum of all entries of  $M$  in two ways.



**3.150.** Draw the functional digraphs associated to each of the following functions  $f : X \rightarrow X$ . For each digraph, find the set  $C$  of cyclic vertices and the set partition  $\{S_v : v \in C\}$  described in 3.43. (a)  $X = \{1, 2, 3, 4\}$ ,  $f$  is the identity map on  $X$ ; (b)  $X = \{0, 1, \dots, 6\}$ ,  $f(x) = (x^2 + 1) \bmod 7$ ; (c)  $X = \{0, 1, \dots, 12\}$ ,  $f(x) = (x^2 + 1) \bmod 13$ ; (d)  $X = \{0, 1, \dots, 10\}$ ,  $f(x) = 3x \bmod 11$ ; (e)  $X = \{0, 1, \dots, 11\}$ ,  $f(x) = 4x \bmod 12$ .

**3.151.** Let  $X = \{0, 1, 2, \dots, 9\}$ . (a) Define  $f : X \rightarrow X$  by setting  $f(x) = (3x + 7) \bmod 10$ . Draw the functional digraphs for  $f$ ,  $f^{-1}$  and  $f \circ f$ . What is the smallest integer  $k > 0$  such that  $f \circ f \circ \dots \circ f$  ( $k$  factors) is the identity map on  $X$ ? (b) Define  $g : X \rightarrow X$  by setting  $g(x) = (2x + 3) \bmod 10$ . Draw the functional digraphs for  $g$  and  $g \circ g$ .

**3.152.** Let  $X$  be a finite set, let  $x_0 \in X$ , and let  $f : X \rightarrow X$  be any function. Recursively define  $x_{m+1} = f(x_m)$  for all  $m \geq 0$ . Show that there exists  $i > 0$  with  $x_i = x_{2i}$ .

**3.153. Pollard-rho Factoring Algorithm.** Suppose  $N > 1$  is an integer. Let  $X = \{0, 1, \dots, N-1\}$ , and define  $f : X \rightarrow X$  by  $f(x) = (x^2 + 1) \bmod N$ . (a) Show that the following algorithm always terminates and returns a divisor of  $N$  greater than 1. (Use 3.152.)

**Step 1.** Set  $u = f(0)$ ,  $v = f(f(0))$ , and  $d = \gcd(v - u, N)$ .

**Step 2.** While  $d = 1$ : set  $u = f(u)$ ,  $v = f(f(v))$ , and  $d = \gcd(v - u, N)$ .

**Step 3.** Return  $d$ .

(b) Trace the steps taken by this algorithm to factor  $N = 77$  and  $N = 527$ .

**3.154.** Suppose  $X$  is a finite set of size  $k$  and  $f : X \rightarrow X$  is a random function (so for all  $x, y \in X$ ,  $P(f(x) = y) = 1/k$ , and these events are independent for different choices of  $x$ ). Let  $x_0 \in X$ , define  $x_{m+1} = f(x_m)$  for all  $m \geq 0$ , and let  $S$  be the least index such that  $x_S = x_t$  for some  $t < S$ . (a) For each  $s \geq 0$ , find the exact probability that  $S > s$ . (b) Argue informally that the expected value of  $S$  is at most  $2\sqrt{k}$ . (c) Use (b) to argue informally that the expected number of gcd computations needed by the Pollard-rho factoring algorithm to find a divisor of a composite number  $N$  (see 3.153) is bounded above by  $2N^{1/4}$ .

**3.155.** Let  $V$  be an  $n$ -element set, and let  $v_0 \notin V$ . A function  $f : V \rightarrow V$  is called *acyclic* iff all cycles in the functional digraph of  $f$  have length 1. Count these functions by setting up a bijection between the set of acyclic functions on  $V$  and the set of rooted trees on  $V \cup \{v_0\}$  with root  $v_0$ .

**3.156.** How many bijections  $f$  on an 8-element set are such that the functional digraph of  $f$  has (a) five cycles; (b) three cycles; (c) one cycle?

**3.157.** Let  $X$  be an  $n$ -element set. Let  $Y$  be the set of all functional digraphs for bijections  $f : X \rightarrow X$ . How many equivalence classes does  $Y$  have under the equivalence relation of graph isomorphism (see 3.128)?

**3.158.** How many functional digraphs with vertex set  $\{1, 2, \dots, n\}$  have  $a_1$  cycles of length 1,  $a_2$  cycles of length 2, etc., where  $\sum_i ia_i = n$ ?

**3.159.** Referring to the proof of 3.47, draw pictures of the set  $A$  of functions, the set  $B$  of trees, and the bijection  $\phi : A \rightarrow B$  when  $n = 4$ .

**3.160.** Compute the rooted tree associated to the function below by the map  $\phi$  in the proof of 3.47.

$$\begin{array}{llllll} f(1) = 1; & f(2) = 19; & f(3) = 8; & f(4) = 30; & f(5) = 5; \\ f(6) = 15; & f(7) = 8; & f(8) = 9; & f(9) = 26; & f(10) = 23; \\ f(11) = 21; & f(12) = 30; & f(13) = 27; & f(14) = 13; & f(15) = 28; \\ f(16) = 16; & f(17) = 13; & f(18) = 23; & f(19) = 25; & f(20) = 11; \\ f(21) = 5; & f(22) = 19; & f(23) = 25; & f(24) = 30; & f(25) = 18; \\ f(26) = 9; & f(27) = 16; & f(28) = 15; & f(29) = 7; & f(30) = 30. \end{array}$$

**3.161.** Compute the function associated to the rooted tree with edge set

$$\{(1, 1), (2, 12), (3, 1), (4, 3), (5, 10), (6, 17), (7, 15), (8, 7), (9, 3), \\ (10, 3), (11, 12), (12, 1), (13, 4), (14, 10), (15, 1), (16, 4), (17, 4)\}$$

by the map  $\phi^{-1}$  in the proof of 3.47.

**3.162.** Formulate a theorem for rooted trees similar to 3.75, and prove it by analyzing the bijection in 3.47.

**3.163.** Let  $G$  be the digraph in Figure 3.2. Use the algorithm in 3.52 to convert the walk

$$W = (1, b, 1, b, 1, a, 3, f, 5, m, 2, n, 5, h, 4, c, 3, f, 5, j, 4, g, 5, m, 2, k, 4)$$

to a path in  $G$  from 1 to 4.

**3.164.** What are the strong components of a functional digraph?

**3.165.** Show that a connected graph  $G$  with  $n$  vertices has  $n$  edges iff  $G$  has exactly one cycle.

**3.166.** Prove that a graph  $G$  is not connected iff there exists an ordering of the vertices of  $G$  for which the adjacency matrix of  $G$  is block-diagonal with at least two diagonal blocks.

**3.167.** Prove 3.60 using 3.58, and again without using 3.58.

**3.168.** How many connected simple graphs have vertex set  $\{1, 2, 3, 4\}$ ?

**3.169.** How many connected simple graphs on the vertex set  $\{1, 2, 3, 4, 5\}$  have exactly five edges?

**3.170. Bipartite Graphs.** A graph  $G$  is called *bipartite* iff there exist two sets  $A$  and  $B$  (called *partite sets* for  $G$ ) such that  $A \cap B = \emptyset$ ,  $A \cup B = V(G)$ , and every edge of  $G$  has one endpoint in  $A$  and one endpoint in  $B$ . (a) Prove that a bipartite graph  $G$  has no cycle of odd length. (b) Prove that a graph  $G$  with no odd-length cycles is bipartite by considering, for each component  $C$  of  $G$ , the length of the shortest path from a fixed vertex  $v_0 \in C$  to the other vertices in  $C$  (cf. 3.134). (c) Prove that a graph  $G$  with no odd-length cycles is bipartite by induction on the number of edges in  $G$ .

**3.171.** How many bipartite simple graphs have partite sets  $A = \{1, 2, \dots, m\}$  and  $B = \{m+1, \dots, m+n\}$ ?

**3.172.** Suppose  $G$  is a  $k$ -regular graph with  $n$  vertices. (a) How many edges are in  $G$ ? (b) If  $k > 0$  and  $G$  is bipartite with partite sets  $A$  and  $B$ , prove that  $|A| = |B|$ .

**3.173.** Fix  $k \geq 2$ . Prove or disprove: there exists a  $k$ -regular bipartite graph  $G$  such that  $G$  has a cut-edge.

**3.174.** Prove that an  $n$ -vertex graph  $G$  in which every vertex has degree at least  $(n-1)/2$  must be connected.

**3.175.** Let  $G$  be a forest with  $n$  vertices and  $k$  connected components. Compute  $\sum_{v \in V(G)} \deg_G(v)$  in terms of  $n$  and  $k$ .

**3.176.** The *arboricity* of a simple graph  $G$ , denoted  $\text{arb}(G)$ , is the least  $n$  such that there exist  $n$  forests  $F_i$  with  $V(G) = \bigcup_{i=1}^n V(F_i)$  and  $E(G) = \bigcup_{i=1}^n E(F_i)$ . Prove that

$$\text{arb}(G) \geq \max_H \left\lceil \frac{|E(H)|}{|V(H)| - 1} \right\rceil,$$

where  $H$  ranges over all induced subgraphs of  $G$  with more than one vertex. (It can be shown that equality holds [99].)

**3.177.** Show that any tree not isomorphic to a path graph  $P_n$  (see 3.124(e)) must have at least three leaves.

**3.178.** Let  $T$  be a tree. Show that  $\deg_T(v)$  is odd for all  $v \in V(T)$  iff for all  $e \in E(T)$ , both connected components of  $(V(T), E(T) \setminus \{e\})$  have an odd number of vertices.

**3.179. Helly Property of Trees.** Suppose  $T, T_1, \dots, T_k$  are trees, each  $T_i$  is a subgraph of  $T$ , and  $V(T_i) \cap V(T_j) \neq \emptyset$  for all  $i, j \leq k$ . Show that  $\bigcap_{i=1}^k V(T_i) \neq \emptyset$ .

**3.180.** Let  $G$  be a tree with leaves  $\{v_1, \dots, v_m\}$ . Let  $H$  be a tree with leaves  $\{w_1, \dots, w_m\}$ . Suppose that, for each  $i$  and  $j$ , the length of the unique path in  $G$  from  $v_i$  to  $v_j$  equals the length of the unique path in  $H$  from  $w_i$  to  $w_j$ . Prove  $G \cong H$ .

**3.181.** For  $1 \leq n \leq 7$ , count the number of isomorphism classes of trees with  $n$  vertices.

**3.182.** (a) How many isomorphism classes of  $n$ -vertex trees have exactly 3 leaves? (b) How many trees with vertex set  $\{1, 2, \dots, n\}$  have exactly 3 leaves?

**3.183.** How many trees with vertex set  $\{1, 2, \dots, n\}$  have exactly  $k$  leaves?

**3.184.** Let  $K_n$  be the complete graph on  $n$  vertices (see 3.124). (a) Give a bijective or probabilistic proof that every edge of  $K_n$  appears in the same number of spanning trees of  $K_n$ . (b) Use Cayley's theorem to count the spanning trees of  $K_n$  that do not use the edge  $\{1, 2\}$ .

**3.185.** Use 3.75 to find the number of trees  $T$  with  $V(T) = \{1, 2, \dots, 8\}$  and  $\deg(T) = [3, 3, 3, 1, 1, 1, 1, 1]$ .

**3.186.** Let  $t_n$  be the number of trees on a given  $n$ -element vertex set. Without using Cayley's theorem, prove the recursion

$$t_n = \sum_{k=1}^{n-1} k \binom{n-2}{k-1} t_k t_{n-k}.$$

**3.187.** (a) Use the pruning bijection to find the word associated to the tree

$$T = (\{0, 1, \dots, 8\}, \{\{1, 5\}, \{2, 8\}, \{3, 7\}, \{7, 0\}, \{6, 2\}, \{4, 7\}, \{5, 4\}, \{2, 4\}\}).$$

(b) Use the inverse of the pruning bijection to find the tree with vertex set  $\{0, 1, \dots, 8\}$  associated to the word 1355173.

**3.188.** Use the inverse of the pruning bijection to find all trees with vertex set  $\{1, 2, \dots, 7\}$  associated to the words in  $\mathcal{R}(11334)$ .

**3.189.** Let  $G$  be the graph with vertex set  $\{\pm 1, \pm 2, \dots, \pm n\}$  and with an edge between  $i$  and  $-j$  for all  $i, j \in \{1, 2, \dots, n\}$ . (a) Show that any spanning tree in  $G$  has at least one positive leaf and at least one negative leaf. (b) Develop an analogue of the pruning map that sets up a bijection between the set of spanning trees of  $G$  and pairs of words  $(u, v)$ , where  $u \in \{1, \dots, n\}^{n-1}$  and  $v \in \{-1, \dots, -n\}^{n-1}$ . Conclude that  $G$  has  $n^{2n-2}$  spanning trees.

**3.190.** (a) How many words in  $\mathcal{R}(0^9 1^1 2^1 3^2 4^1)$  are terms? (b) How many words in  $\mathcal{R}(0^8 1^1 2^3 3^1)$  are lists of  $n$  terms? What is  $n$ ?

**3.191.** Given  $w = 00220000201030$ , use the proof of the cycle lemma 3.90 to find all  $i$  such that the cyclic rotation  $R_i(w)$  is a list of 4 terms.

**3.192.** Consider a product  $x_1 \times x_2 \times \cdots \times x_m$  where the binary operation  $\times$  is not necessarily associative. Define a bijection from the set of complete parenthesizations of this product to the set of terms in  $\mathcal{R}(0^m 2^{m-1})$ . Then use 3.91 to show that the number of such parenthesizations is given by a Catalan number.

**3.193.** Let  $\chi_n(x)$  be the chromatic polynomial for the graph  $C_n$  consisting of  $n$  vertices joined in a cycle. Prove that

$$\chi_n(x) = (x-1)^n + (-1)^n(x-1) \quad (n \geq 2).$$

**3.194.** Find the chromatic polynomials for the graphs in 3.124(a),(b),(c),(d).

**3.195.** Find the chromatic polynomial and chromatic number for the graph  $G_2$  in Figure 3.1.

**3.196.** Find two non-isomorphic simple graphs with the same chromatic polynomial.

**3.197.** A certain department wishes to schedule meetings for a number of committees, whose members are listed in the following table.

Committee	Members
Advisory	Driscoll, Loomis, Lasker
Alumni	Sheffield, Loomis
Colloquium	Johnston, Tchaikovsky, Zorn
Computer	Loomis, Clark, Spade
Graduate	Kennedy, Loomis, Trotter
Merit	Lee, Rotman, Fowler, Sheffield
Personnel	Lasker, Schreier, Tchaikovsky, Trotter
Undergraduate	Jensen, Lasker, Schreier, Trotter, Perkins

(a) What is the minimum number of time slots needed so that all committees could meet with no time conflicts? (b) How many non-conflicting schedules are possible if there are six (distinguishable) time slots available? (c) Repeat (a) and (b), assuming that Zorn becomes a member of the merit committee (and remains a member of the colloquium committee).

**3.198.** Let  $K_n$  be the complete graph on  $n$  vertices (see 3.124). (a) How many subgraphs does  $K_n$  have? (b) How many induced subgraphs does  $K_n$  have?

**3.199.** Prove that a graph  $G$  has at least one spanning tree iff  $G$  is connected.

**3.200.** Fill in the details of the proof of 3.111.

**3.201.** Use the spanning tree recursion 3.108 to find  $\tau(G_1)$  for the graph  $G_1$  in Figure 3.1.

**3.202.** Let  $T_1$  and  $T_2$  be spanning trees of a graph  $G$ .

(a) If  $e_1 \in E(T_1) \sim E(T_2)$ , prove there exists  $e_2 \in E(T_2) \sim E(T_1)$  such that

$$T_3 = (V(G), (E(T_1) \sim \{e_1\}) \cup \{e_2\})$$

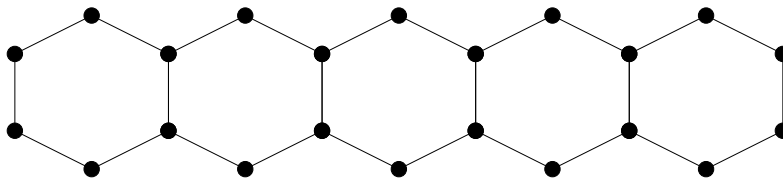
is a spanning tree of  $G$ .

(b) If  $e_1 \in E(T_1) \sim E(T_2)$ , prove there exists  $e_2 \in E(T_2) \sim E(T_1)$  such that

$$T_4 = (V(G), (E(T_2) \cup \{e_1\}) \sim \{e_2\})$$

is a spanning tree of  $G$ .

**3.203.** Fix  $k \geq 3$ . For each  $n \geq 1$ , let  $G_n$  be a graph obtained by gluing together  $n$  regular  $k$ -gons in a row along shared edges. The picture below illustrates the case  $k = 6$ ,  $n = 5$ .



Let  $G_0$  consist of a single edge. Prove the recursion

$$\tau(G_n) = k\tau(G_{n-1}) - \tau(G_{n-2}) \quad (n \geq 2).$$

What are the initial conditions?

**3.204.** Given a simple graph  $G$ , let  $G \sim v$  be the induced subgraph with vertex set  $V(G) \sim \{v\}$ . Assume  $|V(G)| = n \geq 3$ . (a) Prove that  $|E(G)| = (n-2)^{-1} \sum_{v \in V(G)} |E(G \sim v)|$ . (b) Prove that, for  $v_0 \in V(G)$ ,  $\deg_G(v_0) = (n-2)^{-1} \sum_{v \in V(G)} |E(G \sim v)| - |E(G \sim v_0)|$ .

**3.205.** For each graph in 3.124(a) through (f), count the number of spanning trees by direct enumeration, and again by the matrix-tree theorem.

**3.206.** Confirm by direct enumeration that the digraph in Figure 3.25 has 16 spanning trees rooted at 0.

**3.207.** Let  $G$  be the graph with vertex set  $\{0,1\}^3$  such that there is an edge between  $v, w \in V(G)$  iff the words  $v$  and  $w$  differ in exactly one position. Find the number of spanning trees of  $G$ .

**3.208.** Let  $I$  be the  $m \times m$  identity matrix, let  $J$  be the  $m \times m$  matrix all of whose entries are 1, and let  $t, u$  be scalars. Show that  $\det(tI - uJ) = t^m - mt^{m-1}u$ .

**3.209.** Deduce Cayley's theorem 3.72 from the matrix-tree theorem 3.114.

**3.210.** Let  $A$  and  $B$  be disjoint sets of size  $m$  and  $n$ , respectively. Let  $G$  be the simple graph with vertex set  $A \cup B$  and edge set  $\{\{a, b\} : a \in A, b \in B\}$ . Show that  $\tau(G) = m^{n-1}n^{m-1}$ .

**3.211.** How many closed Eulerian tours starting at vertex 5 does the digraph in Figure 3.26 have?

**3.212.** An *Eulerian tour* of a graph  $G$  is a walk in  $G$  that uses every edge exactly once and visits every vertex. (a) Find necessary and sufficient conditions for a graph to have a closed Eulerian tour. (b) Find necessary and sufficient conditions for a graph to have an Eulerian tour.

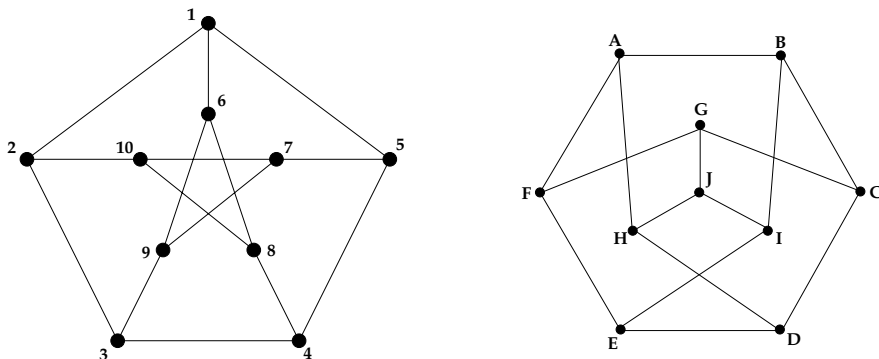
**3.213.** Consider a “digraph with indistinguishable edges” consisting of a vertex set  $V$  and a *multiset* of directed edges  $(u, v) \in V \times V$ . Formulate the notion of a closed Eulerian tour for such a digraph, and prove an analogue of 3.122.

**3.214. de Bruijn Sequences.** Let  $A = \{x_1, \dots, x_n\}$  be an  $n$ -letter alphabet. For each  $k \geq 2$ , show that there exists a word  $w = w_0w_1 \cdots w_{n^k-1}$  such that the  $n^k$  words

$$w_iw_{i+1} \cdots w_{i+k-1} \quad (0 \leq i < n^k)$$

(where subscripts are reduced mod  $n^k$ ) consist of all possible  $k$ -letter words over  $A$ .

**3.215.** The *Petersen graph* is the graph  $G$  with vertex set consisting of all two-element subsets of  $\{1, 2, 3, 4, 5\}$ , and with edge set  $\{\{A, B\} : A \cap B = \emptyset\}$ . (a) Compute the number of vertices and edges in  $G$ . (b) Show that  $G$  is isomorphic to each of the graphs shown here.



(c) Show that  $G$  is 3-regular. (d) Is  $G$  bipartite? (e) Show that any two non-adjacent vertices in  $G$  have exactly one common neighbor.

**3.216.** Find (with proof) all  $k$  such that the Petersen graph has a cycle of length  $k$ .

**3.217.** Given any edge  $e$  in the Petersen graph  $G$ , count the number of cycles of length 5 in  $G$  that contain  $e$ . Use this to count the total number of cycles of length 5 in  $G$ .

**3.218.** (a) Prove that the Petersen graph  $G$  has exactly ten cycles of length 6. (b) How many claws (see 3.124) appear as induced subgraphs of  $G$ ?

**3.219.** How many spanning trees does the Petersen graph have?

## Notes

Our coverage of graph theory in this chapter has been limited to a few enumerative topics. Systematic expositions of graph theory may be found in [14, 17, 18, 27, 59, 67, 136, 143]; the text by West is especially recommended. Roberts [114] gives a treatment of graph theory that emphasizes applications.

The bijection used to enumerate rooted trees in 3.47 is due to Egecioğlu and Remmel [32]. The original proof of Cayley's formula 3.75 appears in Cayley [24]. The pruning bijection described in §3.12 is due to Prüfer [105]; the image of a tree under this map is often called the *Prüfer code* of the tree. For more on the enumeration of trees, see Moon [96].

A version of the cycle lemma 3.90 occurs in the work of Dvoretzky and Motzkin [31]. This lemma and other equivalent results have been independently rediscovered (in various guises) by many authors. Our discussion of the enumeration of lists of terms in §3.14 closely follows Raney's classic paper on Lagrange inversion [107].

The matrix-tree theorem for undirected graphs is usually attributed to Kirchhoff [76]; Tutte extended the theorem to digraphs [132]. The enumeration of Eulerian tours in 3.122 was proved by van Aardenne-Ehrenfest and de Bruijn [133].

This page intentionally left blank

## Inclusion-Exclusion and Related Techniques

This chapter studies combinatorial techniques that are related to the arithmetic operation of subtraction: involutions, inclusion-exclusion formulas, and Möbius inversion. Involutions allow us to give bijective proofs of identities involving both positive and negative terms. The inclusion-exclusion formula extends the sum rule 1.2 to a rule for computing  $|A_1 \cup A_2 \cup \cdots \cup A_m|$  in the case where the sets  $A_i$  need not be pairwise disjoint. This formula turns out to be a special case of the general Möbius inversion formula for posets, which has many applications in number theory and algebra as well as combinatorics.

### 4.1 Involutions

In Chapter 2, we saw how to use bijections to prove combinatorial identities. Many identities involve a mixture of positive and negative terms. One can use *involutions* to furnish combinatorial proofs of such identities. We illustrate the idea using the following binomial coefficient identity.

**4.1. Theorem.** For all  $n \geq 1$ ,  $\sum_{k=0}^n (-1)^k \binom{n}{k} = 0$ .

*Proof.* The result can be proved algebraically by using the binomial theorem 2.14 to expand the left side of  $(-1 + 1)^n = 0$ . To prove the identity combinatorially, let  $X$  be the set of all subsets of  $\{1, 2, \dots, n\}$ . For each  $S \in X$ , we define the *sign* of  $S$  to be  $\text{sgn}(S) = (-1)^{|S|}$ . Since there are  $\binom{n}{k}$  subsets  $S$  of size  $k$ , and  $\text{sgn}(S) = (-1)^k$  for all such subsets, we see that

$$\sum_{S \in X} \text{sgn}(S) = \sum_{k=0}^n (-1)^k \binom{n}{k}.$$

Thus we have found a combinatorial model for the left side of the desired identity, which involves *signed objects*.

Now, define a function  $I : X \rightarrow X$  as follows. Given  $S \in X$ , let  $I(S) = S \cup \{1\}$  if  $1 \notin S$ , and let  $I(S) = S \setminus \{1\}$  if  $1 \in S$ . Observe that  $I(I(S)) = S$  for all  $S \in X$ ; in other words,  $I \circ I = \text{id}_X$ . Thus,  $I$  is a bijection that is equal to its own inverse. Furthermore, since  $|I(S)| = |S| \pm 1$ ,  $\text{sgn}(I(S)) = -\text{sgn}(S)$  for all  $S \in X$ . It follows that  $I$  pairs each positive object in  $X$  with a negative object in  $X$ . Consequently, the number of positive objects in  $X$  equals the number of negative objects in  $X$ , and so  $\sum_{S \in X} \text{sgn}(S) = 0$ .  $\square$

The general setup for involution proofs is described as follows.

**4.2. Definition: Involutions.** An *involution* on a set  $X$  is a function  $I : X \rightarrow X$  such that  $I \circ I = \text{id}_X$ . Equivalently,  $I$  is a bijection on  $X$  and  $I = I^{-1}$ . Given an involution  $I$ , the *fixed point set* of  $I$  is the set  $\text{Fix}(I) = \{x \in X : I(x) = x\}$ , which may be empty. If  $\text{sgn} : X \rightarrow \{+1, -1\}$  is a function that attaches a sign to every object in  $X$ , we say that  $I$  is a *sign-reversing involution* (relative to  $\text{sgn}$ ) iff for all  $x \in X \setminus \text{Fix}(I)$ ,  $\text{sgn}(I(x)) = -\text{sgn}(x)$ .



**4.3. Involution Theorem.** Given a finite set  $X$  of signed objects and a sign-reversing involution  $I$  on  $X$ ,

$$\sum_{x \in X} \text{sgn}(x) = \sum_{x \in \text{Fix}(I)} \text{sgn}(x).$$

*Proof.* Let  $X^+ = \{x \in X \sim \text{Fix}(I) : \text{sgn}(x) = +1\}$  and  $X^- = \{x \in X \sim \text{Fix}(I) : \text{sgn}(x) = -1\}$ . By definition,  $I$  restricts to  $X^+$  and  $X^-$  to give functions  $I^+ : X^+ \rightarrow X^-$  and  $I^- : X^- \rightarrow X^+$  that are mutually inverse bijections. Therefore,  $|X^+| = |X^-|$  and

$$\begin{aligned} \sum_{x \in X} \text{sgn}(x) &= \sum_{x \in X^+} \text{sgn}(x) + \sum_{x \in X^-} \text{sgn}(x) + \sum_{x \in \text{Fix}(I)} \text{sgn}(x) \\ &= |X^+| - |X^-| + \sum_{x \in \text{Fix}(I)} \text{sgn}(x) = \sum_{x \in \text{Fix}(I)} \text{sgn}(x). \quad \square \end{aligned}$$

As a first illustration of the involution theorem, we prove a variation of 4.1.

**4.4. Theorem.** For all  $n \geq 1$ ,

$$\sum_{k=0}^n (-1)^k \binom{2n}{k} = (-1)^n \binom{2n-1}{n}.$$

*Proof.* Let  $X$  be the set of all subsets of  $\{1, 2, \dots, 2n\}$  of size at most  $n$ , and let the sign of a subset  $T$  be  $(-1)^{|T|}$ . The left side of the desired identity is  $\sum_{T \in X} \text{sgn}(T)$ . Next, define an involution  $I$  on  $X$  as follows. If  $T \in X$  and  $1 \in T$ , let  $I(T) = T \sim \{1\}$ . If  $T \in X$  and  $1 \notin T$  and  $|T| < n$ , let  $I(T) = T \cup \{1\}$ . Finally, if  $T \in X$  and  $1 \notin T$  and  $|T| = n$ , let  $I(T) = T$ . One checks immediately that  $I$  is a sign-reversing involution. The fixed points of  $I$  are the  $n$ -element subsets of  $\{1, 2, \dots, 2n\}$  not containing 1. There are  $\binom{2n-1}{n}$  such subsets, and each of them has sign  $(-1)^n$ . So  $\sum_{T \in \text{Fix}(I)} \text{sgn}(T)$  is the right side of the desired identity.  $\square$

**4.5. Theorem.** For all  $n \geq 0$ ,

$$\sum_{k=0}^n (-1)^k \binom{n}{k}^2 = \begin{cases} 0 & \text{if } n \text{ is odd;} \\ (-1)^{n/2} \binom{n}{n/2} & \text{if } n \text{ is even.} \end{cases}$$

*Proof.* Let  $X$  be the set of all pairs  $(S, T)$ , where  $S$  and  $T$  are subsets of  $\{1, 2, \dots, n\}$  of the same size. Define  $\text{sgn}(S, T) = (-1)^{|S|}$ . Then the left side of the desired identity is  $\sum_{(S, T) \in X} \text{sgn}(S, T)$ . We define an involution  $I$  on  $X$  as follows. Given  $(S, T) \in X$ , let  $i$  be the least integer in  $\{1, 2, \dots, n\}$  (if there is one) such that either  $i \notin S$  and  $i \notin T$ , or  $i \in S$  and  $i \in T$ . In the former case, let  $I(S, T) = (S \cup \{i\}, T \cup \{i\})$ ; in the latter case, let  $I(S, T) = (S \sim \{i\}, T \sim \{i\})$ ; if no such  $i$  exists, let  $I(S, T) = (S, T)$ . It is routine to check that  $I$  is a sign-reversing involution; in particular, the designated integer  $i$  in the definition of  $I(S, T)$  is the same as the  $i$  used to calculate  $I(I(S, T))$ . By the involution theorem,

$$\sum_{k=0}^n (-1)^k \binom{n}{k}^2 = \sum_{(S, T) \in \text{Fix}(I)} (-1)^{|S|}.$$

Note that  $(S, T) \in \text{Fix}(I)$  iff for every  $i \leq n$ ,  $i$  lies in exactly one of the two sets  $S$  or  $T$ . This can only happen if  $n$  is even and  $|S| = |T| = n/2$  and  $S = \{1, 2, \dots, n\} \sim T$ . So the fixed point set is empty if  $n$  is odd. If  $n$  is even, we can construct an arbitrary element of  $\text{Fix}(I)$  by choosing any subset  $S$  of size  $n/2$  and letting  $T$  be the complementary subset of  $\{1, 2, \dots, n\}$ . Since there are  $\binom{n}{n/2}$  choices for  $S$ , each with sign  $(-1)^{n/2}$ , the formula in the theorem is proved.  $\square$

**4.6. Example: Stirling Numbers.** Recall that  $s(n, k) = (-1)^{n-k}c(n, k)$ , where  $c(n, k)$  is the number of permutations of an  $n$ -element set whose functional digraph consists of  $k$  cycles (§3.6). We will show that

$$\sum_{k=1}^n s(n, k) = \chi(n = 1) \quad (n \geq 1).$$

Both sides are 1 when  $n = 1$ , so assume  $n > 1$ . Let  $X$  be the set of all permutations of  $\{1, 2, \dots, n\}$ . If  $w \in X$  is a permutation with  $k$  cycles, define  $\text{sgn}(w) = (-1)^k$ . Now  $\sum_{w \in X} \text{sgn}(w) = (-1)^n \sum_{k=1}^n s(n, k)$ , so it suffices to define a sign-reversing involution  $I$  on  $X$  with no fixed points. Given  $w \in X$ , the numbers 1 and 2 either appear in the same cycle of  $w$  or in different cycles. If 1 and 2 are in the same cycle, let the elements on this cycle (starting at 1) be

$$(1, x_1, x_2, \dots, x_k, 2, y_1, y_2, \dots, y_j),$$

where  $j, k \geq 0$ . Define  $I(w)$  by replacing this cycle by the two cycles

$$(1, x_1, x_2, \dots, x_k)(2, y_1, y_2, \dots, y_j)$$

and leaving all other cycles the same. Similarly, if 1 and 2 are in different cycles of  $w$ , write these cycles as

$$(1, x_1, x_2, \dots, x_k)(2, y_1, y_2, \dots, y_j)$$

and define  $I(w)$  by replacing these two cycles by the single cycle

$$(1, x_1, x_2, \dots, x_k, 2, y_1, y_2, \dots, y_j).$$

It is immediate that  $I \circ I = \text{id}_X$ ,  $I$  is sign-reversing, and  $I$  has no fixed points.

We can modify the preceding involution to obtain a combinatorial proof of the identity

$$\sum_{k \geq 0} s(i, k)S(k, j) = \chi(i = j),$$

which we proved algebraically in part (d) of 2.77. If  $i < j$ , then for every  $k$ , either  $s(i, k) = 0$  or  $S(k, j) = 0$ . So both sides of the identity are zero in this case. If  $i = j$ , the left side reduces to  $s(i, i)S(i, i) = 1 = \chi(i = j)$ . If  $j = 0$ , the identity is true. So we may assume  $i$  and  $j$  are fixed numbers such that  $i > j > 0$ . Let  $X$  be the set of pairs  $(w, U)$ , where  $w$  is a permutation of  $\{1, 2, \dots, i\}$  (viewed as a functional digraph) and  $U$  is a set partition of the set of cycles in  $w$  into  $j$  blocks. If  $w$  has  $k$  cycles, let  $\text{sgn}(w, U) = (-1)^k$ . Then

$$\sum_{(w, U) \in X} \text{sgn}(w, U) = (-1)^i \sum_{k=j}^i s(i, k)S(k, j)$$

and  $\chi(i = j) = 0$ . So it suffices to define a sign-reversing involution  $I$  on  $X$  with no fixed points. Given  $(w, U) \in X$ , there must exist a block of  $U$  such that the cycles in this block collectively involve more than one point in  $\{1, 2, \dots, i\}$ . This follows from the fact that  $i$  (the number of points) exceeds  $j$  (the number of blocks). Among all such blocks in  $U$ , choose the block that contains the smallest possible element in  $\{1, 2, \dots, i\}$ . Let this smallest element be  $a$ , and let the second-smallest element in this block be  $b$ . To calculate  $I(w, U)$ , modify the cycles in this block of  $U$  as we did above, with  $a$  and  $b$  playing the roles of 1 and 2. More specifically, a cycle of the form

$$(a, x_1, \dots, x_k, b, y_1, \dots, y_j)$$

gets replaced (within its block) by

$$(a, x_1, \dots, x_k)(b, y_1, \dots, y_j)$$

and vice versa. It is routine to check that  $I$  is a sign-reversing involution on  $X$  with no fixed points. For example, suppose  $i = 10$ ,  $j = 3$ ,  $w$  has cycles  $(1), (3, 5), (2, 6, 9), (4, 8), (7), (10)$ , and

$$U = \{\{(1)\}, \{(3, 5), (10)\}, \{(2, 6, 9), (4, 8), (7)\}\}.$$

Here the block of  $U$  modified by the involution is  $\{(2, 6, 9), (4, 8), (7)\}$ ,  $a = 2$ , and  $b = 4$ . We compute  $I(w, U)$  by replacing the cycles  $(2, 6, 9)$  and  $(4, 8)$  in  $w$  by the single cycle  $(2, 6, 9, 4, 8)$  and letting the new set partition be

$$\{\{(1)\}, \{(3, 5), (10)\}, \{(2, 6, 9, 4, 8), (7)\}\}.$$

## 4.2 The Inclusion-Exclusion Formula

Recall the sum rule: if  $S_1, \dots, S_n$  are *pairwise disjoint* finite sets, then  $|S_1 \cup \dots \cup S_n| = \sum_{i=1}^n |S_i|$ . Can we find a formula for  $|S_1 \cup \dots \cup S_n|$  in the case where the given sets  $S_i$  are not necessarily disjoint? The answer is provided by the inclusion-exclusion formula, which we discuss now.

We have already seen the simplest case of the inclusion-exclusion formula. Specifically, if  $S$  and  $T$  are any two finite sets, the binary union rule 1.4 states that

$$|S \cup T| = |S| + |T| - |S \cap T|.$$

Intuitively, the sum  $|S| + |T|$  overestimates the cardinality of  $|S \cup T|$  because elements of  $|S \cap T|$  are included twice in this sum. To correct this, we exclude one copy of each of the elements in  $S \cap T$  by subtracting  $|S \cap T|$ .

Now consider three finite sets  $S$ ,  $T$ , and  $U$ . The sum  $|S| + |T| + |U|$  overcounts the size of  $|S \cup T \cup U|$  since elements in the overlaps between these sets are counted twice (or three times, in the case of elements  $z \in S \cap T \cap U$ ). We may try to account for this by subtracting  $|S \cap T| + |S \cap U| + |T \cap U|$  from  $|S| + |T| + |U|$ . If  $x$  belongs to  $S$  and  $U$  but not  $T$  (say), this subtraction will cause  $x$  to be counted only once in the overall expression. A similar comment applies to elements in  $(S \cap T) \sim U$  and  $(T \cap U) \sim S$ . However, an element  $z \in S \cap T \cap U$  is counted three times in  $|S| + |T| + |U|$  and subtracted three times in  $|S \cap T| + |S \cap U| + |T \cap U|$ . So we must include such elements once again by adding the term  $|S \cap T \cap U|$ . In summary, we have given an informal argument suggesting that the formula

$$|S \cup T \cup U| = |S| + |T| + |U| - |S \cap T| - |S \cap U| - |T \cap U| + |S \cap T \cap U|$$

should be true.

Generalizing the pattern in the preceding example, we arrive at the following formula, known as the *inclusion-exclusion formula*.

**4.7. Inclusion-Exclusion Formula.** Suppose  $n > 0$  and  $S_1, \dots, S_n$  are any finite sets. Then

$$|S_1 \cup S_2 \cup \dots \cup S_n| = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} |S_{i_1} \cap S_{i_2} \cap \dots \cap S_{i_k}|. \quad (4.1)$$

**4.8. Example.** If  $n = 4$ , the inclusion-exclusion formula for  $|S_1 \cup S_2 \cup S_3 \cup S_4|$  is

$$\begin{aligned} & |S_1| + |S_2| + |S_3| + |S_4| \\ & - |S_1 \cap S_2| - |S_1 \cap S_3| - |S_1 \cap S_4| - |S_2 \cap S_3| - |S_2 \cap S_4| - |S_3 \cap S_4| \\ & + |S_1 \cap S_2 \cap S_3| + |S_1 \cap S_2 \cap S_4| + |S_1 \cap S_3 \cap S_4| + |S_2 \cap S_3 \cap S_4| \\ & - |S_1 \cap S_2 \cap S_3 \cap S_4|. \end{aligned}$$

**4.9. Remark.** By setting  $I = \{i_1, i_2, \dots, i_k\}$ , the inclusion-exclusion formula can also be written

$$|S_1 \cup \dots \cup S_n| = \sum_{\emptyset \neq I \subseteq \{1, 2, \dots, n\}} (-1)^{|I|-1} \left| \bigcap_{i \in I} S_i \right|.$$

We will give several proofs of the inclusion-exclusion formula. Each proof illustrates different techniques and can be generalized in different ways.

**4.10. Proof of Inclusion-Exclusion by Induction.** We prove that (4.1) holds for all  $n > 0$  and all finite sets  $S_1, \dots, S_n$  by induction on  $n$ . The formula reduces to  $|S_1| = |S_1|$  for  $n = 1$ , and this is certainly true. For  $n = 2$ , the formula becomes

$$|S_1 \cup S_2| = |S_1| + |S_2| - |S_1 \cap S_2|,$$

and this is the binary union rule 1.4 proved previously. Now assume  $n > 2$  and that formula (4.1) is already known to hold for any union of  $n - 1$  finite sets. Let  $S_1, \dots, S_n$  be fixed finite sets. The  $n$ -fold union  $S_1 \cup \dots \cup S_n$  can be regarded as the union of the two sets  $S = S_1 \cup S_2 \cup \dots \cup S_{n-1}$  and  $T = S_n$ . Hence, by the binary union rule 1.4,

$$|S_1 \cup \dots \cup S_n| = |S_1 \cup \dots \cup S_{n-1}| + |S_n| - |(S_1 \cup \dots \cup S_{n-1}) \cap S_n|.$$

Since the set operations  $\cap$  and  $\cup$  obey the distributive law, we can write the subtracted term as

$$|(S_1 \cap S_n) \cup (S_2 \cap S_n) \cup \dots \cup (S_{n-1} \cap S_n)|,$$

which is the union of the  $n - 1$  finite sets  $S_i \cap S_n$  ( $1 \leq i \leq n - 1$ ). So we can apply the induction hypothesis to this term, and to the first term  $|S_1 \cup \dots \cup S_{n-1}|$ . We obtain

$$\begin{aligned} |S_1 \cup \dots \cup S_n| &= |S_n| + \sum_{k=1}^{n-1} (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n-1} |S_{i_1} \cap \dots \cap S_{i_k}| \\ &\quad - \sum_{j=1}^{n-1} (-1)^{j-1} \sum_{1 \leq i_1 < \dots < i_j \leq n-1} |(S_{i_1} \cap S_n) \cap \dots \cap (S_{i_j} \cap S_n)|. \end{aligned}$$

We modify the second line of this formula as follows. First, observe that

$$\bigcap_{r=1}^j (S_{i_r} \cap S_n) = S_{i_1} \cap S_{i_2} \cap \dots \cap S_{i_j} \cap S_n.$$

Next, change the summation index by setting  $k = j + 1$  and defining  $i_k = n$ . The formula now reads

$$\begin{aligned} |S_1 \cup \dots \cup S_n| &= \sum_{k=1}^{n-1} (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n-1} |S_{i_1} \cap \dots \cap S_{i_k}| \\ &\quad + |S_n| + \sum_{k=2}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_{k-1} < i_k = n} |S_{i_1} \cap \dots \cap S_{i_k}|. \end{aligned}$$

We can absorb  $|S_n|$  into the sum on the second line by allowing  $k$  to range from 1 to  $n$  there. Also, letting  $k$  range from 1 to  $n$  in the first summation does not introduce any new terms. After making these adjustments, the only difference between the formulas on the first and second lines is that  $i_k < n$  in the first line while  $i_k = n$  in the second line. We can now combine the two summations to obtain

$$|S_1 \cup \cdots \cup S_n| = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} |S_{i_1} \cap \cdots \cap S_{i_k}|, \quad (4.2)$$

which is the desired formula (4.1). This completes the induction.

In some counting problems, the following versions of the inclusion-exclusion formula are needed.

**4.11. Alternate Version of Inclusion-Exclusion Formula.** Suppose  $S_1, \dots, S_n$  are subsets of a finite set  $X$ . The number of elements  $x \in X$  that belong to *none* of the  $S_i$  is

$$|X \sim (S_1 \cup \cdots \cup S_n)| = |X| + \sum_{k=1}^n (-1)^k \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} |S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_k}|.$$

This formula follows from the original inclusion-exclusion formula and the difference rule 1.3.

Intuitively, the preceding formula is applicable when we are trying to count objects in  $X$  that must simultaneously avoid a number of specified “bad” properties. Each set  $S_i$  consists of those objects in  $X$  that have the  $i$ th bad property (and possibly other bad properties too).

**4.12. Simplified Version of the Inclusion-Exclusion Formula.** Let  $S_1, \dots, S_n$  be finite sets. Suppose that for all  $k \geq 1$ , the intersection of any  $k$  distinct sets among the  $S_j$ ’s always has cardinality  $N(k)$ . In other words,  $|S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_k}| = N(k)$  for all choices of  $i_1 < i_2 < \cdots < i_k$ . Then

$$|S_1 \cup \cdots \cup S_n| = \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} N(k).$$

If all  $S_j$ ’s are subsets of a given finite set  $X$ , we also have

$$|X \sim (S_1 \cup \cdots \cup S_n)| = |X| + \sum_{k=1}^n (-1)^k \binom{n}{k} N(k).$$

These formulas follow by substituting  $N(k)$  for each summand  $|S_{i_1} \cap \cdots \cap S_{i_k}|$  in the previous inclusion-exclusion formulas and noting that there are  $\binom{n}{k}$  such summands.

### 4.3 More Proofs of Inclusion-Exclusion

This section presents two proofs of the inclusion-exclusion formula that are more combinatorial than the inductive computation already given.

**4.13. Involution Proof of Inclusion-Exclusion.** If we move all terms in (4.1) to the left side, we obtain the formula

$$\sum_{k=0}^n (-1)^k \sum_{1 \leq i_1 < \dots < i_k \leq n} |S_{i_1} \cap \dots \cap S_{i_k}| = 0. \quad (4.3)$$

In this equation, the summand corresponding to  $k = 0$  is defined to be  $|S_1 \cup \dots \cup S_n|$ . We will prove this formula by introducing an involution on a suitable set of signed objects.

Let  $X$  be the set of all sequences  $(x; i_1, i_2, \dots, i_k)$  such that  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ ,  $0 \leq k \leq n$ , and  $x \in S_{i_1} \cap \dots \cap S_{i_k}$ . (If  $k = 0$ , then the object looks like  $(x; )$ , and the last condition is interpreted to mean  $x \in S_1 \cup \dots \cup S_n$ .) Define  $\text{sgn}(x; i_1, i_2, \dots, i_k) = (-1)^k$ . It follows from the sum rule that  $\sum_{z \in X} \text{sgn}(z)$  is the left side of (4.3). So it suffices to define a sign-reversing involution on  $X$  with no fixed points.

Given  $z = (x; i_1, \dots, i_k) \in X$ , we must have  $x \in S_1 \cup S_2 \cup \dots \cup S_n$  no matter what the value of  $k$  is. Let  $i$  be the minimum index in  $\{1, 2, \dots, n\}$  such that  $x \in S_i$ . By definition of  $X$ , we either have  $k = 0$  or  $i < i_1$  or  $i = i_1$ . If  $k = 0$  or  $i < i_1$ , define  $I(z) = (x; i, i_1, i_2, \dots, i_k)$ . If instead  $i = i_1$ , define  $I(z) = (x; i_2, \dots, i_k)$ . It is immediate that  $I(I(z)) = z$  and  $\text{sgn}(I(z)) = -\text{sgn}(z)$  for all  $z \in X$ .

The preceding proof is quite ingenious, since it establishes a rather complicated formula by a remarkably simple bookkeeping bijection. On the other hand, it would be nice to have a combinatorial proof of inclusion-exclusion that is tied more closely to the intuitive “including and excluding” arguments we used originally to guess the formula for  $|S \cup T \cup U|$ . We present such a proof next.

**4.14. Counting Proof of Inclusion-Exclusion.** Fix  $n$  finite sets  $S_1, \dots, S_n$ , and put  $X = S_1 \cup \dots \cup S_n$ . We consider a large matrix  $A$  whose rows are indexed by the elements  $x \in X$  and whose columns are indexed by all nonempty subsets  $T$  of  $\{1, 2, \dots, n\}$ . Define the entry in row  $x$  and column  $T$  of  $A$  to be  $(-1)^{|T|-1}$  iff  $x \in \bigcap_{i \in T} S_i$ , and define this entry to be zero otherwise. The sum of the entries in the column of  $A$  indexed by  $T = \{i_1 < i_2 < \dots < i_k\} \subseteq \{1, 2, \dots, n\}$  is

$$(-1)^{k-1} |S_{i_1} \cap \dots \cap S_{i_k}|.$$

Adding up all these column sums, we see that the sum  $s$  of all entries in  $A$  is

$$s = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} |S_{i_1} \cap \dots \cap S_{i_k}|.$$

Now, let us compute  $s$  by adding up the row sums of  $A$ . Intuitively, the sum of the 1's and  $-1$ 's in row  $x$  of  $A$  represents the net number of times  $x$  has been counted in the inclusion-exclusion sum written above. We claim that this number is 1 for all  $x \in X$ , so that the sum of the row sums is  $s = |X| = |S_1 \cup \dots \cup S_n|$ . This will complete the proof of the inclusion-exclusion formula.

Fix  $x \in X$ , and let  $U = \{i_1 < i_2 < \dots < i_m\}$  be the set of all indices  $i_j$  such that  $x \in S_{i_j}$ . We have  $m > 0$  since  $x$  lies in at least one  $S_i$ . The entry in row  $x$  and column  $T$  of  $A$  is  $(-1)^{|T|-1}$  if  $T \subseteq U$ , and this entry is zero if  $T$  is not a subset of  $U$ . Note that there are  $\binom{m}{k}$  subsets of  $U$  of size  $k$ , each of which contributes  $(-1)^{k-1}$  to the row sum. Grouping all such terms together and invoking the binomial theorem, we conclude that the sum of the entries in row  $x$  of  $A$  is

$$\sum_{k=1}^m \binom{m}{k} (-1)^{k-1} = 1 - \sum_{k=0}^m \binom{m}{k} (-1)^k 1^{m-k} = 1 - (-1 + 1)^m = 1.$$

#### 4.4 Applications of the Inclusion-Exclusion Formula

We can use the inclusion-exclusion formula to count complicated combinatorial collections that cannot be conveniently enumerated by the sum and product rules alone. Recall that, when using inclusion-exclusion, we often set up the problem so that each set  $S_i$  consists of those objects in some big set  $X$  that have a certain “bad” property. Our desired answer is then the cardinality of  $X \sim (S_1 \cup \cdots \cup S_n)$ , which is given by the inclusion-exclusion formula in 4.11.

**4.15. Example: Bridge Hands.** A *bridge hand* is a 13-element subset of a 52-card deck. A *face card* is a jack, queen, king, or ace. How many bridge hands have at least one of each kind of face card? To answer this question, let  $X$  be the set of all bridge hands; note  $|X| = \binom{52}{13}$ . Define  $S_1$  (resp.  $S_2, S_3, S_4$ ) to be the set of all hands in  $X$  that do *not* have a jack (resp. queen, king, ace). The card hands we want are the elements of the set  $X \sim (S_1 \cup S_2 \cup S_3 \cup S_4)$ . We must now compute the sizes of the various intersections  $S_{i_1} \cap \cdots \cap S_{i_k}$ . Note that  $|S_1| = \binom{48}{13}$  since we can build all hands in  $S_1$  by choosing 13 cards out of the 48 non-jacks in the deck. Similarly,  $|S_2| = |S_3| = |S_4| = \binom{48}{13}$ . Next,  $|S_1 \cap S_3| = \binom{44}{13}$  since we can build hands in  $S_1 \cap S_3$  by choosing 13 cards out of the 44 cards in the deck that are neither jacks nor kings. The same formula holds for all other twofold intersections. Similarly, each threefold intersection has size  $\binom{40}{13}$ , while  $|S_1 \cap S_2 \cap S_3 \cap S_4| = \binom{36}{13}$ . It follows from inclusion-exclusion that the answer to the original question is

$$\binom{52}{13} - 4\binom{48}{13} + 6\binom{44}{13} - 4\binom{40}{13} + \binom{36}{13} = 128,971,619,088.$$

Next, how many 13-card bridge hands have at least one jack, at least one queen, and at least one king, but do not contain any ace cards or spade cards? The last condition can be dealt with as follows: throw out the  $13 + 4 - 1 = 16$  aces and spades at the outset, leaving  $52 - 16 = 36$  cards. An inclusion-exclusion argument like the one in the last paragraph now leads to the answer

$$\binom{36}{13} - 3\binom{33}{13} + 3\binom{30}{13} - \binom{27}{13} = 930,511,530.$$

**4.16. Example.** How many words  $w \in X = \mathcal{R}(1^2 2^2 3^2 \cdots n^2)$  never have two adjacent letters that are equal? Note first that  $|X| = \binom{2n}{2, 2, \dots, 2} = (2n)!/2^n$ . For  $1 \leq i \leq n$ , let  $S_i$  be the set of words in  $X$  in which the two copies of letter  $i$  are adjacent to each other. We wish to count the words in  $X \sim (S_1 \cup \cdots \cup S_n)$ . To do so, fix  $i_1 < i_2 < \cdots < i_k$  and consider a typical intersection  $S_{i_1} \cap \cdots \cap S_{i_k}$ . Given a word  $w$  in this intersection, form a new word by replacing the two consecutive copies of  $i_j$  by a single copy of  $i_j$ , for  $1 \leq j \leq k$ . This operation defines a bijection from  $S_{i_1} \cap \cdots \cap S_{i_k}$  onto the set  $\mathcal{R}(1^{a_1} 2^{a_2} \cdots n^{a_n})$ , where  $a_i = 1$  if  $i = i_j$  for some  $j$ , and  $a_i = 2$  otherwise. (The inverse bijection replaces each  $i_j$  by two consecutive copies of  $i_j$ .) It follows that

$$|S_{i_1} \cap \cdots \cap S_{i_k}| = \binom{1k + 2(n-k)}{\underbrace{1, \dots, 1}_k, \underbrace{2, \dots, 2}_{n-k}} = (2n-k)!/2^{n-k}.$$

This expression does not depend on the indices  $i_1, \dots, i_k$ . Also, when  $k = 0$ , this expression reduces to  $|X|$ . Using the simplified inclusion-exclusion formula 4.12, we conclude that

$$|X \sim (S_1 \cup \cdots \cup S_n)| = \sum_{k=0}^n (-1)^k \binom{n}{k} \frac{(2n-k)!}{2^{n-k}}.$$

For our next examples, we use inclusion-exclusion to enumerate certain combinatorial collections that have arisen in earlier chapters.

**4.17. Theorem: Enumeration of Surjections.** Let  $\text{Surj}(m, n)$  be the number of surjections from an  $m$ -element set onto an  $n$ -element set. If  $m \geq n \geq 1$ , then

$$\text{Surj}(m, n) = \sum_{k=0}^n (-1)^k \binom{n}{k} (n-k)^m.$$

*Proof.* Let  $X$  be the set of all functions  $f : \{1, 2, \dots, m\} \rightarrow \{1, 2, \dots, n\}$ . Note that  $|X| = n^m$ . For  $1 \leq i \leq n$ , let  $S_i$  consist of all functions  $f \in X$  such that  $i$  is *not* in the image of  $f$ . A function  $f \in X$  is a surjection iff  $f$  belongs to none of the  $S_i$ . Thus, we must compute  $|X \sim (S_1 \cup \dots \cup S_n)|$ . Consider a typical intersection  $S_{i_1} \cap \dots \cap S_{i_k}$ , where  $i_1 < i_2 < \dots < i_k$ . A function  $f$  belonging to this intersection is the same thing as an arbitrary function mapping  $\{1, 2, \dots, m\}$  into the  $(n-k)$ -element set  $\{1, 2, \dots, n\} \sim \{i_1, i_2, \dots, i_k\}$ . The number of such functions is  $(n-k)^m$ , independent of  $i_1, \dots, i_k$ . Using the simplified inclusion-exclusion formula 4.12, we get

$$\text{Surj}(m, n) = |X \sim (S_1 \cup \dots \cup S_n)| = n^m + \sum_{k=1}^n (-1)^k \binom{n}{k} (n-k)^m,$$

which is equivalent to the formula of the theorem.  $\square$

Since  $S(m, n) = \text{Surj}(m, n)/n!$  by 2.58, we deduce the following formula for Stirling numbers of the second kind.

**4.18. Theorem: Summation Formula for Stirling Numbers of the Second Kind.**

$$S(m, n) = \frac{1}{n!} \sum_{k=0}^n (-1)^k \binom{n}{k} (n-k)^m = \sum_{k=0}^n (-1)^k \frac{(n-k)^m}{k!(n-k)!}.$$

Our next illustration of inclusion-exclusion comes from number theory.

**4.19. Definition: Euler's  $\phi$  Function.** For each integer  $m \geq 1$ , let  $\phi(m)$  be the number of integers  $x \in \{1, 2, \dots, m\}$  such that  $\gcd(x, m) = 1$ .

For example, if  $m = 12$ , then the relevant integers  $x$  are 1, 5, 7, and 11, so  $\phi(12) = 4$ . The function  $\phi$  is prominent in algebra and number theory and has applications to modern cryptography.

**4.20. Theorem: Formula for  $\phi(m)$ .** Suppose an integer  $m > 1$  has prime factorization  $m = p_1^{e_1} p_2^{e_2} \dots p_n^{e_n}$ . Then

$$\phi(m) = \prod_{i=1}^n p_i^{e_i-1} (p_i - 1) = m \prod_{i=1}^n (1 - 1/p_i).$$

*Proof.* Let  $X = \{1, 2, \dots, m\}$ , and let  $S_i = \{x \in X : p_i | x\}$ . (The symbol  $p_i | x$  means that  $p_i$  divides  $x$ .) By the fundamental theorem of arithmetic,  $x \in X$  is not relatively prime to  $m$  iff  $x$  and  $m$  have a common factor greater than 1 iff  $x$  and  $m$  have a common *prime* factor. It follows that

$$\phi(m) = |X \sim (S_1 \cup S_2 \cup \dots \cup S_n)|.$$



So we are in a position to use inclusion-exclusion. Here it is convenient to write the inclusion-exclusion formula as follows:

$$|X \sim (S_1 \cup \cdots \cup S_n)| = \sum_{I \subseteq \{1, 2, \dots, n\}} (-1)^{|I|} \left| \bigcap_{i \in I} S_i \right|,$$

where we interpret  $\bigcap_{i \in \emptyset} S_i$  as the set  $X$ . Fix a subset  $I = \{i_1 < \cdots < i_k\} \subseteq \{1, 2, \dots, n\}$ , and consider the intersection  $S_{i_1} \cap \cdots \cap S_{i_k}$ . An integer  $x \leq m$  lies in this intersection iff  $p_{i_j} | x$  for  $1 \leq j \leq k$  iff the product  $q = p_{i_1} p_{i_2} \cdots p_{i_k}$  divides  $x$  iff  $x$  is a multiple of  $q$ . Now, the number of multiples of  $q$  between 1 and  $m$  is  $m/q = m / \prod_{i \in I} p_i$ . If  $I = \emptyset$  and the empty product is interpreted as 1, this expression becomes  $m = |X|$ . Hence, the inclusion-exclusion formula can be written

$$\phi(m) = |X \sim (S_1 \cup \cdots \cup S_n)| = m \sum_{I \subseteq \{1, 2, \dots, n\}} (-1)^{|I|} \frac{1}{\prod_{i \in I} p_i}.$$

On the other hand, consider what happens when we expand the product

$$m \prod_{i=1}^n \left(1 - \frac{1}{p_i}\right)$$

using the generalized distributive law (cf. 2.7). We will obtain a sum of  $2^n$  terms, each of which is obtained by choosing either 1 or  $-\frac{1}{p_i}$  from the  $i$ th factor of the product and multiplying these choices together. We can index these  $2^n$  terms by subsets  $I \subseteq \{1, 2, \dots, n\}$ , where  $i \in I$  iff we chose  $-\frac{1}{p_i}$  from the  $i$ th factor. It follows that

$$m \prod_{i=1}^n \left(1 - \frac{1}{p_i}\right) = m \sum_{I \subseteq \{1, 2, \dots, n\}} (-1)^{|I|} \frac{1}{\prod_{i \in I} p_i} = \phi(m). \quad \square$$

**4.21. Remark.** We sketch an alternative proof of the formula for  $\phi(m)$  that avoids inclusion-exclusion. This proof sketch will use some facts from algebra and number theory without proof. For any commutative ring  $R$ , we let  $R^\times$  be the set of *units* in  $R$ ; i.e., the set of  $x \in R$  such that there exists  $y \in R$  with  $xy = yx = 1_R$ . The following facts are routinely verified. First, if  $R$  and  $S$  are isomorphic rings, then  $|R^\times| = |S^\times|$ . Second, given a product ring  $R \times S$ , we have  $(R \times S)^\times = R^\times \times S^\times$  and hence (by the product rule)  $|(R \times S)^\times| = |R^\times| \cdot |S^\times|$ . Third,  $\gcd(x, n) = 1$  iff there exist integers  $y, z$  with  $xy + nz = 1$  iff  $x$  has a multiplicative inverse in the ring of integers modulo  $n$ . So  $\phi(n) = |(\mathbb{Z}/n\mathbb{Z})^\times|$ . Fourth, by the Chinese Remainder Theorem, the rings  $\mathbb{Z}/mn\mathbb{Z}$  and  $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$  are isomorphic whenever  $\gcd(m, n) = 1$ . Combining these four facts, we see that  $\gcd(m, n) = 1$  implies

$$\phi(mn) = |(\mathbb{Z}/mn\mathbb{Z})^\times| = |(\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z})^\times| = |(\mathbb{Z}/m\mathbb{Z})^\times| \cdot |(\mathbb{Z}/n\mathbb{Z})^\times| = \phi(m)\phi(n).$$

Iteration of this result gives

$$\phi(p_1^{e_1} \cdots p_n^{e_n}) = \prod_{i=1}^n \phi(p_i^{e_i})$$

whenever  $p_1, \dots, p_n$  are distinct primes. Thus, it suffices to evaluate  $\phi$  at prime powers. But a direct counting argument using the difference rule and the definition of  $\phi$  shows that  $\phi(p^e) = p^e - p^{e-1} = p^{e-1}(p - 1)$  when  $p$  is prime and  $e \geq 1$ . So we obtain the first formula for  $\phi(n)$  given in 4.20.

## 4.5 Derangements

The inclusion-exclusion formula allows us to enumerate a special class of permutations called derangements. Intuitively, a derangement of  $1, 2, \dots, n$  is a rearrangement of these  $n$  symbols such that no symbol remains in its original position. The formal definition is as follows.

**4.22. Definition: Derangements.** A *derangement* of a set  $S$  is a bijection  $f : S \rightarrow S$  such that  $f(x) \neq x$  for all  $x \in S$ . For  $n \geq 0$ , let  $D_n$  be the set of derangements of  $\{1, 2, \dots, n\}$ , and let  $d_n = |D_n|$ .

Note that  $d_0 = 1$  (since the function with empty graph satisfies the definition of derangement), while  $d_1 = 0$ . To give more examples of derangements, let us identify an element  $f \in D_n$  with the word  $f(1)f(2)\cdots f(n)$ . Then  $d_2 = 1$  since 21 is the unique derangement of two letters. The derangements of three letters are 312 and 231, so that  $d_3 = 2$ . The permutation 5317426 is a derangement of seven letters.

**4.23. Summation Formula for Derangements.** For  $n \geq 1$ , the number of derangements of  $n$  letters is

$$d_n = n! \sum_{k=0}^n (-1)^k \frac{1}{k!}.$$

Consequently,  $d_n$  is the closest integer to  $n!/e$  for  $n \geq 1$ .

*Proof.* Let  $X$  be the set of all permutations of  $n$  letters; note that  $|X| = n!$ . For  $1 \leq i \leq n$ , let  $S_i = \{f \in X : f(i) = i\}$ . The set  $D_n$  consists of precisely those elements in  $X$  that belong to none of the  $S_i$ , so  $D_n = X \setminus (S_1 \cup \cdots \cup S_n)$ . To apply the inclusion-exclusion formula, we must consider a typical intersection  $S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_k}$ , where  $i_1 < i_2 < \cdots < i_k$ . A permutation  $f \in X$  belongs to this intersection iff  $f$  fixes  $i_1, \dots, i_k$  and permutes the remaining  $n - k$  letters among themselves. The number of such permutations is  $(n - k)!$ . This number depends only on  $k$  and not on the indices  $i_1, \dots, i_k$ . Applying the simplified inclusion-exclusion formula 4.12, we obtain

$$d_n = n! + \sum_{k=1}^n (-1)^k \binom{n}{k} (n - k)! = n! + \sum_{k=1}^n (-1)^k \frac{n!}{k!} = n! \sum_{k=0}^n (-1)^k \frac{1}{k!}.$$

To relate this formula to the expression  $n!/e$ , recall from calculus that

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (x \in \mathbb{R}).$$

Setting  $x = -1$ , we see that

$$1/e = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!}.$$

Multiplying by  $n!$  and comparing to our formula for  $d_n$ , we see that

$$n!/e - d_n = n! \sum_{k=n+1}^{\infty} \frac{(-1)^k}{k!}.$$

It now suffices to show that the right side of this formula is less than  $1/2$  in absolute value. Factoring out  $\frac{1}{(n+1)!}$  from each term in the series, we obtain

$$|n!/e - d_n| = \frac{1}{n+1} \left| 1 - \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} - \frac{1}{(n+2)(n+3)(n+4)} + \cdots \right|.$$

The series within the absolute values on the right side is an alternating series that converges to a sum strictly less than 1. Since  $n \geq 1$ , it follows that

$$|n!/e - d_n| < \frac{1}{n+1} \cdot 1 \leq 1/2. \quad \square$$

The following table lists the first few values of  $d_n$ .

$n$	0	1	2	3	4	5	6	7	8	9
$d_n$	1	0	1	2	9	44	265	1854	14,833	133,496

Like any permutation, a derangement has a functional digraph consisting of the disjoint union of one or more cycles. A permutation is a derangement iff there are no 1-cycles in its functional digraph. This observation leads to the following recursion for derangements.

**4.24. Theorem: Recursion for Derangements.** We have  $d_0 = 1$ ,  $d_1 = 0$ , and

$$d_n = (n-1)d_{n-1} + (n-1)d_{n-2} \quad (n \geq 2).$$

*Proof.* Fix  $n \geq 2$ . Write the set of derangements  $D_n$  as the disjoint union of sets  $A$  and  $B$ , where  $A$  consists of those derangements in which  $n$  is involved in a cycle of length 2, and  $B$  consists of the derangements where  $n$  is in a cycle of length greater than 2. To build an object in  $A$ , choose the partner of  $n$  in its 2-cycle ( $n-1$  ways), and then choose a derangement of the remaining objects ( $D_{n-2}$  ways). To build an object in  $B$ , choose a derangement of the first  $n-1$  objects ( $D_{n-1}$  ways), consider the functional digraph of this derangement, and splice  $n$  into a cycle just before any of the  $n-1$  available elements (which is guaranteed to create a cycle of length 3 or more). The recursion now follows from the sum and product rules.  $\square$

**4.25. Theorem: Second Recursion for Derangements.** We have  $d_0 = 1$  and

$$d_n = nd_{n-1} + (-1)^n \quad (n \geq 1).$$

*Proof.* We argue by induction on  $n$ . If  $n = 1$ , then

$$d_n = d_1 = 0 = 1 \cdot 1 + (-1)^1 = nd_{n-1} + (-1)^n.$$

Now assume  $n > 1$  and that  $d_{n-1} = (n-1)d_{n-2} + (-1)^{n-1}$ . We can use this assumption to eliminate  $(n-1)d_{n-2}$  in the first recursion 4.24 for  $d_n$  (which is already known to hold for all  $n$ ). We thereby obtain

$$d_n = (n-1)d_{n-1} + (n-1)d_{n-2} = (n-1)d_{n-1} + (d_{n-1} - (-1)^{n-1}) = nd_{n-1} + (-1)^n.$$

This completes the induction.  $\square$

## 4.6 Coefficients of Chromatic Polynomials

Let  $G$  be a simple graph. Recall that  $\chi_G(x)$  denotes the number of proper colorings of the vertices of  $G$  using  $x$  available colors. We have seen in 3.100 that  $\chi_G(x)$  is always a *polynomial* in  $x$ . In this section, we use inclusion-exclusion to analyze the chromatic polynomial of  $G$ . This analysis will lead to a combinatorial interpretation for the coefficients of the chromatic polynomial  $\chi_G(x)$ .

**4.26. Definition: Vertex-spanning Subgraph.** Let  $G = (V(G), E(G))$  be a simple graph. A *vertex-spanning subgraph* of  $G$  is a subgraph  $H$  of  $G$  such that  $V(H) = V(G)$ .

The map  $H \mapsto E(H)$  is a bijection between the set of vertex-spanning subgraphs of  $G$  and the set of all subsets of  $E(G)$ .

**4.27. Theorem: Coefficients of Chromatic Polynomials.** Let  $G$  be a simple graph. For each  $e, c \geq 0$ , let  $n(e, c)$  be the number of vertex-spanning subgraphs of  $G$  with  $e$  edges and  $c$  connected components. Then

$$\chi_G(x) = \sum_{e, c \geq 0} (-1)^e n(e, c) x^c.$$

*Proof.* Let  $e_1, \dots, e_n$  be the edges of  $G$ . Let  $X$  be the set of all colorings of  $G$  (proper or not) using  $x$  available colors, and let  $S_i$  be the set of colorings in  $X$  such that both endpoints of the edge  $e_i$  receive the same color. We wish to compute  $|X \sim (S_1 \cup \dots \cup S_n)|$ . Consider a typical intersection  $\bigcap_{i \in T} S_i$ , where  $T \subseteq \{1, 2, \dots, n\}$ . The edge subset  $\{e_i : i \in T\}$  determines a vertex-spanning subgraph  $H$  of  $G$  with  $|T|$  edges and some number  $cc(H)$  of connected components. One may check that a coloring  $f$  belongs to  $\bigcap_{i \in T} S_i$  iff  $f$  is constant on each connected component of  $H$ . It follows from the product rule that  $|\bigcap_{i \in T} S_i| = x^{cc(H)}$ , since we can choose one of  $x$  colors for each connected component of  $H$ . Note also that  $|X| = x^{|V(G)|} = x^{cc(H_0)}$  where  $H_0 = (V(G), \emptyset)$ . By inclusion-exclusion,

$$\begin{aligned} \chi_G(x) &= |X| + \sum_{\emptyset \neq T \subseteq \{1, 2, \dots, n\}} (-1)^{|T|} \left| \bigcap_{i \in T} S_i \right| \\ &= \sum_{\text{vertex-spanning subgraphs } H} (-1)^{|E(H)|} x^{cc(H)} = \sum_{e, c \geq 0} (-1)^e n(e, c) x^c. \quad \square \end{aligned}$$

## 4.7 Classical Möbius Inversion

We conclude this chapter with a brief introduction to the theory of Möbius inversion. We begin in this section by studying the number-theoretic Möbius function and the classical Möbius inversion formula. Later sections discuss the generalization of the Möbius function and inversion formula to posets.

**4.28. Definition: Classical Möbius Function.** Suppose  $m \geq 1$  is an integer with prime factorization  $m = p_1^{e_1} p_2^{e_2} \dots p_n^{e_n}$ , where  $n \geq 0$ ,  $e_i > 0$ , and the  $p_i$ 's are distinct primes. (We take  $n = 0$  when  $m = 1$ .) The *Möbius function*  $\mu : \mathbb{N}^+ \rightarrow \{-1, 0, 1\}$  is defined by  $\mu(m) = 0$  if  $e_i > 1$  for some  $i$ , whereas  $\mu(m) = (-1)^n$  if  $e_i = 1$  for all  $i$ .

In other words,  $\mu(m)$  is zero if  $m$  is divisible by the square of a prime;  $\mu(m) = +1$  if  $m$  is the product of an even number of distinct primes; and  $\mu(m) = -1$  if  $m$  is the product of an odd number of distinct primes. For example,

$$\mu(1) = 1, \mu(7) = -1, \mu(10) = 1, \mu(12) = 0, \mu(30) = -1.$$

The following theorem is the key to proving the Möbius inversion formula.

**4.29. Theorem.** For all integers  $m \geq 1$ ,  $\sum_{d|m} \mu(d) = \chi(m = 1)$ . (Here and below, the symbol  $\sum_{d|m}$  means that we sum over all *positive* divisors  $d$  of the integer  $m$ .)

*Proof.* When  $m = 1$ , we have  $\sum_{d|1} \mu(d) = \mu(1) = 1 = \chi(m = 1)$ . Suppose next that  $m > 1$  and  $m$  has prime factorization  $p_1^{e_1} \cdots p_n^{e_n}$ . Instead of summing  $\mu(d)$  over *all* divisors  $d$  of  $m$ , we may equally well sum over just the *square-free* divisors  $d$  of  $m$ , which give the only nonzero contributions to the sum. Examining prime factorizations, we see that there are  $2^n$  such square-free divisors, which have the form  $\prod_{i \in T} p_i$  as  $T$  ranges over all subsets of  $\{1, 2, \dots, n\}$ . Therefore,

$$\sum_{d|m} \mu(d) = \sum_{T \subseteq \{1, 2, \dots, n\}} \mu\left(\prod_{i \in T} p_i\right) = \sum_{T \subseteq \{1, 2, \dots, n\}} (-1)^{|T|}.$$

Collecting together summands indexed by subsets  $T$  of the same size  $k$ , we conclude that

$$\sum_{d|m} \mu(d) = \sum_{k=0}^n \sum_{\substack{T \subseteq \{1, \dots, n\} \\ |T|=k}} (-1)^{|T|} = \sum_{k=0}^n \binom{n}{k} (-1)^k 1^{n-k} = (-1 + 1)^n = 0. \quad \square$$

**4.30. Classical Möbius Inversion Formula.** Suppose  $f$  and  $g$  are functions with domain  $\mathbb{N}^+$  such that

$$f(m) = \sum_{d|m} g(d) \quad (m \geq 1).$$

Then

$$g(m) = \sum_{d|m} f(m/d) \mu(d) = \sum_{d|m} f(d) \mu(m/d) \quad (m \geq 1).$$

*Proof.* We use the definition of  $f$  to expand the first claimed formula for  $g(m)$ :

$$\sum_{d|m} f(m/d) \mu(d) = \sum_{d|m} \left( \sum_{c|(m/d)} g(c) \right) \mu(d) = \sum_{(c,d) \in S} g(c) \mu(d),$$

where  $S = \{(c, d) \in \mathbb{N}^+ \times \mathbb{N}^+ : d|m, c|(m/d)\}$ . It follows routinely from the definition of divisibility that

$$S = \{(c, d) : d|m, cd|m\} = \{(c, d) : c|m, cd|m\} = \{(c, d) : c|m, d|(m/c)\}.$$

Therefore, the calculation continues as follows:

$$\begin{aligned} \sum_{(c,d) \in S} g(c) \mu(d) &= \sum_{c|m} \sum_{d|(m/c)} g(c) \mu(d) = \sum_{c|m} g(c) \left( \sum_{d|(m/c)} \mu(d) \right) \\ &= \sum_{c|m} g(c) \chi(m/c = 1) = g(m). \end{aligned}$$

The next-to-last step used 4.29 to simplify the inner sum. We conclude that

$$g(m) = \sum_{d|m} f(m/d)\mu(d) = \sum_{d|m} f(d)\mu(m/d),$$

where the final equality results by replacing  $d$  by  $m/d$  in the summation.  $\square$

To give examples of the Möbius inversion formula, we first introduce some functions that are studied in number theory.

**4.31. Definition: Number-Theoretic Functions  $\tau$ ,  $\sigma$ , and  $\sigma_2$ .** Let  $m \geq 1$  be an integer. Define

$$\tau(m) = \sum_{d|m} 1; \quad \sigma(m) = \sum_{d|m} d; \quad \sigma_2(m) = \sum_{d|m} d^2.$$

Thus,  $\tau(m)$  is the number of positive divisors of  $m$ ;  $\sigma(m)$  is the sum of these divisors; and  $\sigma_2(m)$  is the sum of the squares of these divisors.

**4.32. Example.** Taking  $m = 1, 4, 7, 12, 30$ , we calculate:

$$\begin{array}{llllll} \tau(1) = 1, & \tau(4) = 3, & \tau(7) = 2, & \tau(12) = 6, & \tau(30) = 8; \\ \sigma(1) = 1, & \sigma(4) = 7, & \sigma(7) = 8, & \sigma(12) = 28, & \sigma(30) = 72; \\ \sigma_2(1) = 1, & \sigma_2(4) = 21, & \sigma_2(7) = 50, & \sigma_2(12) = 210, & \sigma_2(30) = 1300. \end{array}$$

If  $m$  has prime factorization  $p_1^{e_1} \cdots p_n^{e_n}$ , then the divisors of  $m$  have the form  $p_1^{f_1} \cdots p_n^{f_n}$  where  $0 \leq f_i \leq e_i$  for all  $i$ . The product rule therefore gives  $\tau(m) = \prod_{i=1}^n (e_i + 1)$  (build a divisor by choosing  $f_1, \dots, f_n$ ). Using the generalized distributive law and the geometric series formula, one may also check that

$$\sigma(m) = \prod_{i=1}^n \left( \sum_{f_i=0}^{e_i} p_i^{f_i} \right) = \prod_{i=1}^n \frac{p_i^{e_i+1} - 1}{p_i - 1}.$$

Applying the Möbius inversion formula to the definitions of  $\tau$ ,  $\sigma$ , and  $\sigma_2$ , we obtain the following identities.

**4.33. Theorem.** For  $m \geq 1$ , we have

$$1 = \sum_{d|m} \tau(m/d)\mu(d); \quad m = \sum_{d|m} \sigma(m/d)\mu(d); \quad m^2 = \sum_{d|m} \sigma_2(m/d)\mu(d).$$

The next result uses Möbius inversion to deduce information about Euler's  $\phi$  function.

**4.34. Theorem:  $\phi$  versus  $\mu$ .** For all  $m \geq 1$ ,

$$m = \sum_{d|m} \phi(d) \quad \text{and so} \quad \phi(m) = \sum_{d|m} \mu(d)(m/d).$$

*Proof.* To prove the first formula, fix  $m \geq 1$ . For each divisor  $d$  of  $m$ , let

$$S_d = \{x \in \mathbb{N}^+ : 1 \leq x \leq m \text{ and } \gcd(x, m) = d\}.$$

It is immediate that the  $m$ -element set  $\{1, 2, \dots, m\}$  is the disjoint union of the sets  $S_d$  as  $d$  ranges over the positive divisors of  $m$ . Whenever  $d$  divides  $m$ , we have  $\gcd(x, m) = d$  iff  $d$  divides  $x$  and  $\gcd(x/d, m/d) = 1$ . It follows that division by  $d$  gives a bijection from the

set  $S_d$  onto the set of numbers counted by  $\phi(m/d)$ . Therefore,  $|S_d| = \phi(m/d)$ . By the sum rule,

$$m = \sum_{d|m} |S_d| = \sum_{d|m} \phi(m/d) = \sum_{d|m} \phi(d).$$

The last equality follows by noting that the number  $m/d$  ranges over all positive divisors of  $m$  as  $d$  ranges over all positive divisors of  $m$ . Applying Möbius inversion (with  $f(m) = m$  and  $g(m) = \phi(m)$ ), we obtain the second formula in the theorem.  $\square$

Some applications of these results to field theory are presented in §12.6.

## 4.8 Partially Ordered Sets

We will see that the inclusion-exclusion formula 4.7 and the classical Möbius inversion formula 4.30 are special cases of the general Möbius inversion formula for partially ordered sets (posets). First we must review some definitions and examples concerning posets.

Recall from 2.54 the definition of relations and the notions of reflexive, irreflexive, symmetric, antisymmetric, and transitive relations. Given a relation  $R$  on a finite set  $X$ , the pair  $(X, R)$  is a digraph  $G$  with vertex set  $X$  and directed edge set  $R$ . Reflexivity means that *every* vertex of  $G$  has a loop edge; irreflexivity means that *no* vertex of  $G$  has a loop edge. Symmetry means that the reversal of every edge is also an edge (so we can think of  $G$  as undirected); antisymmetry means that it is never true that a non-loop edge and its reversal are both in  $G$ . Finally, transitivity means that whenever there is a walk  $(x, y, z)$  of length 2 in  $G$ , the edge  $(x, z)$  is also present in  $G$ . More generally, we see by induction that when  $R$  is transitive, there exists a walk from  $x$  to  $z$  in  $G$  of positive length iff the edge  $(x, z)$  is present in  $G$ .

**4.35. Poset Definitions.** A *partial order relation* on  $X$  is a relation that is antisymmetric, transitive, and reflexive on  $X$ . A *strict order relation* on  $X$  is a relation that is transitive and irreflexive on  $X$ . A *partially ordered set (poset)* is a pair  $(X, \leq)$  where  $\leq$  is a partial order relation on  $X$ . A *totally ordered set* is a poset  $(X, \leq)$  such that for all  $x, y \in X$ , either  $x \leq y$  or  $y \leq x$ .

**4.36. Example.** Let  $X = \{1, 2, \dots, n\}$  and take  $\leq$  to be the usual ordering of integers. Then  $(X, \leq)$  is an  $n$ -element totally ordered poset. More generally, for any  $S \subseteq \mathbb{R}$ ,  $(S, \leq)$  is a totally ordered poset.

**4.37. Example: Boolean Posets.** Let  $S$  be any set, and let  $X = \mathcal{P}(S)$  be the set of all subsets of  $S$ . Then  $(X, \subseteq)$  is a poset, where  $A \subseteq B$  means that  $A$  is a subset of  $B$ . In particular,  $(\mathcal{P}(\{1, 2, \dots, n\}), \subseteq)$  is a poset of size  $2^n$ . This poset is not totally ordered when  $n > 1$ .

**4.38. Example: Divisibility Posets.** Consider the divisibility relation  $|$  on  $\mathbb{N}^+$  defined by  $a|b$  iff  $b = ac$  for some  $c \in \mathbb{N}^+$ . Then  $(\mathbb{N}^+, |)$  is an infinite poset. Given a fixed positive integer  $n$ , let  $X$  be the set of all divisors of  $n$ . Restricting  $|$  to  $X$  gives a finite poset  $(X, |)$ . This poset is a totally ordered set iff  $n$  is a prime power.

The next result shows that partial order relations and strict order relations are essentially equivalent concepts.

**4.39. Theorem: Partial Orders vs. Strict Orders.** Let  $X$  be a set, let  $P$  be the set of all partial order relations on  $X$ , and let  $S$  be the set of all strict order relations on  $X$ . There is a canonical bijection between  $P$  and  $S$ .

*Proof.* Let  $\Delta = \{(x, x) : x \in X\}$  be the “diagonal” of  $X \times X$ . Define  $f : P \rightarrow S$  by setting  $f(\leq) = \leq \sim \Delta$  for each partial ordering  $\leq$  on  $X$ . Define  $g : S \rightarrow P$  by setting  $g(<) = < \cup \Delta$  for each strict ordering  $<$  on  $X$ . In terms of the digraphs,  $f$  removes self-loops from all vertices and  $g$  restores the self-loops. It is an exercise for the reader to show that  $f$  does map  $P$  into  $S$ ,  $g$  does map  $S$  into  $P$ , and  $f \circ g$  and  $g \circ f$  are both identity maps.  $\square$

## 4.9 Möbius Inversion for Posets

**4.40. Definition: Matrix of a Relation.** Let  $X = \{x_1, x_2, \dots, x_n\}$  be a finite set, and let  $R$  be a relation on  $X$ . Define the *matrix of  $R$*  to be the  $n \times n$  matrix  $A = A(R)$  such that  $A_{i,j} = \chi(x_i R x_j)$ .  $A(R)$  is the adjacency matrix of the digraph  $(X, R)$ .

**4.41. Theorem.** Let  $\leq$  be a partial ordering of  $X = \{x_1, \dots, x_n\}$ , and let  $<$  be the associated strict ordering of  $X$  (see 4.39). Consider the matrices  $Z = A(\leq)$  and  $N = A(<)$ . Then  $Z = I + N$ ;  $N$  is nilpotent;  $Z$  is invertible; and

$$Z^{-1} = I - N + N^2 - N^3 + \dots + (-1)^{n-1} N^{n-1}. \quad (4.4)$$

*Proof.* The matrix identity  $Z = I + N$  holds since  $(X, \leq)$  is obtained from  $(X, <)$  by adding self-loops at each  $x \in X$ . Next, we claim that the digraph  $(X, <)$  is acyclic. For if  $(z_1, z_2, \dots, z_k, z_1)$  were a directed cycle in this digraph, we must have  $z_1 < z_2 < \dots < z_k < z_1$ . Then transitivity gives  $z_1 < z_1$ , which contradicts irreflexivity. By 3.24,  $N$  is nilpotent. The statements about the inverse of  $Z$  now follow from 3.25, taking  $A$  there to be  $-N$ .  $\square$

**4.42. Definition: Möbius Function of a Finite Poset.** Keeping the notation of the preceding theorem, define  $\mu = \mu_{(X, \leq)} : X \times X \rightarrow \mathbb{Z}$  by setting  $\mu(x_i, x_j)$  to be the  $i, j$ -entry of  $Z^{-1}$ . The function  $\mu$  is called the *Möbius function of the poset  $(X, \leq)$* .

**4.43. Example.** Let  $X = \{1, 2, 3, 4\}$  with the usual ordering. For this poset, we have

$$Z = A(\leq) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad N = A(<) = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The powers of  $N$  are

$$N^2 = \begin{pmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad N^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad N^4 = 0.$$

So the inverse of  $Z$  is

$$Z^{-1} = I - N + N^2 - N^3 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

So  $\mu(i, i) = 1$ ,  $\mu(i, i+1) = -1$ , and  $\mu(i, j) = 0$  for all  $j \neq i, i+1$ .



**4.44. Example: Möbius Function of a Totally Ordered Poset.** The preceding example generalizes as follows. Let  $X = \{1, 2, \dots, n\}$  with the usual ordering. We have  $Z_{i,j} = 1$  for  $i \leq j$  and  $Z_{i,j} = 0$  for  $i > j$ . For all  $i$ , let  $M_{i,i} = 1$ ,  $M_{i,i+1} = -1$ , and  $M_{i,j} = 0$  for  $j \neq i, i+1$ . A routine matrix calculation shows that  $ZM = MZ = I$ . So for this poset,

$$\mu(i, i) = 1, \quad \mu(i, i+1) = -1, \quad \mu(i, j) = 0 \text{ for } j \neq i, i+1.$$

**4.45. Example: Möbius Function for Boolean Posets.** Consider the poset  $(X, \subseteq)$ , where  $X$  consists of all subsets of  $[n] = \{1, 2, \dots, n\}$ . In this example, we will index the rows and columns of matrices by subsets of  $[n]$ . For  $S, T \subseteq [n]$ , the  $S, T$ -entry of  $Z$  is 1 if  $S \subseteq T$ , and 0 otherwise. We claim that the inverse matrix  $M = Z^{-1}$  has  $S, T$ -entry  $\mu(S, T) = (-1)^{|T \setminus S|}$  if  $S \subseteq T$ , and zero otherwise. To verify this, let us show that  $ZM = I$ . The  $S, T$ -entry of  $ZM$  is

$$\sum_{U \subseteq [n]} Z(S, U)M(U, T) = \sum_{U: S \subseteq U \subseteq T} (-1)^{|T \setminus U|}.$$

If  $S = T$ , this sum is 1; while if  $S \not\subseteq T$ , this sum is 0. Now consider the case where  $S \subsetneq T$ . Let  $S$  have  $a$  elements and  $T$  have  $a + b$  elements, where  $b > 0$ . For  $0 \leq c \leq b$ , the number of sets  $U$  with  $S \subseteq U \subseteq T$  and  $|T \setminus U| = c$  is  $\binom{b}{c}$ . Grouping terms in the sum based on the size of  $|T \setminus U|$ , we see that

$$(ZM)(S, T) = \sum_{c=0}^b (-1)^c \binom{b}{c} = (-1 + 1)^b = 0.$$

So the Möbius function for this poset is

$$\mu(S, T) = (-1)^{|T \setminus S|} \chi(S \subseteq T) \quad (S, T \subseteq [n]).$$

An alternate proof of this formula will be given in 4.58 below.

**4.46. Example: Möbius Function for Divisibility Posets.** Let  $n$  be a fixed positive integer, let  $X$  be the set of positive divisors of  $n$ , and consider the divisibility poset  $(X, |)$ . There is a close relation between the classical Möbius function  $\mu$  and the Möbius function  $\mu_X$  for this poset. More precisely, we claim that

$$\mu(d) = \mu_X(1, d) \quad \text{for all } d \text{ dividing } n.$$

To verify this, let us work with matrices whose rows and columns are indexed by the positive divisors of  $n$ , considered in increasing order. As above, let  $Z$  be the matrix such that  $Z_{d,e} = \chi(d|e)$ ; let  $M$  be the inverse matrix, which is uniquely determined by  $Z$ ; and let  $\vec{v}$  be the row vector  $(\mu(d) : d|n)$ . The identity  $\sum_{d|m} \mu(d) = \chi(m=1)$ , which is valid for all  $m$  dividing  $n$ , can now be rewritten as the vector identity  $\vec{v}Z = (1, 0, \dots, 0)$ . This shows that  $\vec{v}$  must be the first row of  $M$ . It will be shown in 4.59 that  $\mu_X(d, e) = \mu(e/d)$  whenever  $d|e$  and  $e|n$ , whereas  $\mu_X(d, e) = 0$  if  $d$  does not divide  $e$ .

The next definition will be used to give a combinatorial interpretation for the values of the Möbius function.

**4.47. Definition: Chains in a Poset.** Let  $(X, \leq)$  be a poset. A *chain of length  $k$*  in  $X$  is a sequence  $C = (z_0, z_1, \dots, z_k)$  of elements of  $X$  such that

$$z_0 < z_1 < \dots < z_k.$$

We say that  $C$  is a chain *from  $z_0$  to  $z_k$*  and write  $\text{len}(C) = k$ . The *sign* of the chain  $C$  is  $\text{sgn}(C) = (-1)^k$ .

**4.48. Theorem: Möbius Functions and Signed Chains.** Let  $(X, \leq)$  be a finite poset. Given  $y, z \in X$ , let  $S$  be the set of all chains in  $X$  from  $y$  to  $z$ . Then

$$\mu_{(X, \leq)}(y, z) = \sum_{C \in S} \text{sgn}(C).$$

In particular, if  $y \not\leq z$ , then  $\mu_{(X, \leq)}(y, z) = 0$ .

*Proof.* We know from (4.4) that

$$\mu_{(X, \leq)}(y, z) = \sum_{k \geq 0} (-1)^k N^k(y, z),$$

where  $N$  is the adjacency matrix of the digraph  $G = (X, <)$ . A chain of length  $k$  from  $y$  to  $z$  is the same as a walk (or path) of length  $k$  from  $y$  to  $z$  in  $G$ . By 3.18, the number of such walks is  $N^k(y, z)$ . The theorem now follows from the sum rule.  $\square$

**4.49. Theorem: Möbius Inversion Formula on Posets.** Let  $(X, \leq)$  be a finite poset with Möbius function  $\mu$ . Suppose  $R$  is a commutative ring and  $f, g : X \rightarrow R$  are two functions. Then

$$\left( \forall x \in X, g(x) = \sum_{y \leq x} f(y) \right) \quad \text{iff} \quad \left( \forall x \in X, f(x) = \sum_{y \leq x} g(y) \mu(y, x) \right).$$

*Proof.* Let  $X = \{x_1, \dots, x_n\}$ , and define  $Z = A(\leq)$  and  $M = Z^{-1}$  as in 4.41. Also define row vectors  $F = [f(x_1), \dots, f(x_n)]$  and  $G = [g(x_1), \dots, g(x_n)]$ . The left-hand formula in the theorem is equivalent to the matrix identity  $G = FZ$ , since  $G_j = g(x_j)$  and

$$(FZ)_j = \sum_{k=1}^n F_k Z_{k,j} = \sum_{k=1}^n f(x_k) \chi(x_k \leq x_j) = \sum_{y \leq x_j} f(y).$$

Similarly, keeping in mind that  $\mu(y, x) = 0$  unless  $y \leq x$ , the right-hand formula in the theorem is equivalent to the matrix identity  $F = GM$ . Since  $M$  and  $Z$  are inverse matrices,  $G = FZ$  is equivalent to  $GM = F$ .  $\square$

**4.50. Example.** In the special case where  $X = \{1, 2, \dots, n\}$  with the usual ordering, 4.49 reduces to the following statement: given  $f_1, \dots, f_n \in R$  and  $g_1, \dots, g_n \in R$ , we have  $(g_i = f_1 + f_2 + \dots + f_i \text{ for all } i)$  iff  $(f_1 = g_1 \text{ and } f_i = g_i - g_{i-1} \text{ for } 1 < i \leq n)$ .

**4.51. Example.** In the special case where  $X$  is the set of divisors of  $n$  ordered by divisibility, 4.49 reduces to the classical inversion formula 4.30, using the fact that  $\mu_X(d, e) = \mu(e/d)$  when  $d|e$ , and  $\mu_X(d, e) = 0$  otherwise.

**4.52. Example.** In the special case where  $X = \mathcal{P}([n])$  ordered by containment of subsets, 4.49 reduces to the following statement:

$$\left( \forall T \subseteq [n], g(T) = \sum_{S \subseteq T} f(S) \right) \quad \text{iff} \quad \left( \forall T \subseteq [n], f(T) = \sum_{S \subseteq T} (-1)^{|T \sim S|} g(S) \right).$$

If instead we use the “opposite” poset  $(X, \supseteq)$ , one obtains:

$$\left( \forall T \subseteq [n], g(T) = \sum_{S \supseteq T} f(S) \right) \quad \text{iff} \quad \left( \forall T \subseteq [n], f(T) = \sum_{S \supseteq T} (-1)^{|S \sim T|} g(S) \right).$$

We now use this result to rederive a version of the original inclusion-exclusion formula. Let  $Z_1, \dots, Z_n$  be given subsets of a finite set  $Z$ . For  $S \subseteq [n]$ , let  $f(S)$  be the number of objects  $z \in Z$  such that  $z \in Z_i$  if and only if  $i \in S$ . For  $S \subseteq [n]$ , let  $g(S)$  be the number of objects  $z \in Z$  such that  $z \in Z_i$  if  $i \in S$ . Regarding  $Z_i$  as the set of objects in  $Z$  with a certain property  $i$ , we can say that  $f(S)$  counts objects that have *exactly* the properties in  $S$ , whereas  $g(S)$  counts the objects that have *at least* the properties in  $S$ . It follows from this that  $g(T) = \sum_{S \supseteq T} f(S)$  for all  $T$ , so 4.49 tells us that  $f(T) = \sum_{S \supseteq T} (-1)^{|S \setminus T|} g(S)$  for all  $T$ . Now,  $f(\emptyset) = |\bar{Z}| \sim (Z_1 \cup \dots \cup Z_n)$  and  $g(\{i_1, \dots, i_k\}) = |Z_{i_1} \cap \dots \cap Z_{i_k}|$ . The formula in 4.11 follows from these observations.

## 4.10 Product Posets

This section introduces a construction on posets that leads to alternative derivations of the Möbius functions for the posets  $(\mathcal{P}([n]), \subseteq)$  and  $(\{d : d|n\}, |)$ .

**4.53. Definition: Product Posets.** Let  $(X_1, \leq_1), \dots, (X_n, \leq_n)$  be posets. Consider the Cartesian product  $X = X_1 \times \dots \times X_n$ , which consists of all  $n$ -tuples  $x = (x_1, \dots, x_n)$  with  $x_i \in X_i$ . For  $x = (x_i)$  and  $y = (y_i)$  in  $X$ , define  $x \leq y$  iff  $x_i \leq_i y_i$  for  $1 \leq i \leq n$ . One immediately verifies that  $\leq$  is a partial ordering on  $X$ . The poset  $(X, \leq)$  is called the *product* of the posets  $(X_i, \leq_i)$ .

**4.54. Example.** Let  $X_1 = X_2 = \{1, 2\}$  with the usual ordering. Both  $X_1$  and  $X_2$  are totally ordered posets, but  $X = X_1 \times X_2$  is not totally ordered. For example,  $(1, 2)$  and  $(2, 1)$  are two incomparable elements of  $X$ .

**4.55. Theorem: Möbius Function for a Product Poset.** Let  $(X, \leq)$  be the product of posets  $(X_i, \leq_i)$  for  $1 \leq i \leq k$ . Given  $x = (x_i)$  and  $y = (y_i)$  in  $X$ , we have

$$\mu_{(X, \leq)}(x, y) = \prod_{i=1}^k \mu_{(X_i, \leq_i)}(x_i, y_i).$$

*Proof.* For brevity, write  $\mu = \mu_{(X, \leq)}$  and  $\mu_i = \mu_{(X_i, \leq_i)}$ . By induction, we can reduce to the case  $k = 2$ . We have the matrices

$$\begin{aligned} Z_1 &= [\chi(u_1 \leq_1 v_1) : u_1, v_1 \in X_1], & M_1 &= [\mu_1(u_1, v_1) : u_1, v_1 \in X_1], \\ Z_2 &= [\chi(u_2 \leq_2 v_2) : u_2, v_2 \in X_2], & M_2 &= [\mu_2(u_2, v_2) : u_2, v_2 \in X_2], \\ Z &= [\chi(u \leq v) : u, v \in X], & M &= [\mu(u, v) : u, v \in X], \end{aligned}$$

which satisfy  $Z_1 M_1 = I$ ,  $Z_2 M_2 = I$  and  $ZM = I$ . Define a matrix  $M'$ , with rows and columns indexed by elements of  $X$ , such that for  $u = (u_1, u_2)$  and  $v = (v_1, v_2)$  in  $X$ , the  $u, v$ -entry of  $M'$  is  $\mu_1(u_1, v_1)\mu_2(u_2, v_2)$ . Note that the  $u, v$ -entry of  $Z$  is  $\chi((u_1, u_2) \leq (v_1, v_2)) = \chi(u_1 \leq_1 v_1)\chi(u_2 \leq_2 v_2)$ . The following computation verifies that  $ZM' = I$ , and hence  $M' = M$ :

$$\begin{aligned} (ZM')(u, w) &= \sum_{v \in X} Z(u, v)M'(v, w) \\ &= \sum_{v_1 \in X_1} \sum_{v_2 \in X_2} \chi(u_1 \leq_1 v_1)\chi(u_2 \leq_2 v_2)\mu_1(v_1, w_1)\mu_2(v_2, w_2) \\ &= \left( \sum_{v_1 \in X_1} \chi(u_1 \leq_1 v_1)\mu_1(v_1, w_1) \right) \cdot \left( \sum_{v_2 \in X_2} \chi(u_2 \leq_2 v_2)\mu_2(v_2, w_2) \right) \\ &= \chi(u_1 = w_1)\chi(u_2 = w_2) = \chi(u = w). \quad \square \end{aligned}$$

**4.56. Definition: Poset Isomorphisms.** Given posets  $(X, \leq)$  and  $(X', \leq')$ , a *poset isomorphism* is a bijection  $f : X \rightarrow X'$  such that

$$u \leq v \text{ iff } f(u) \leq' f(v) \quad (u, v \in X).$$

**4.57. Theorem.** If  $f : (X, \leq) \rightarrow (X', \leq')$  is a poset isomorphism, then

$$\mu_{(X', \leq')} (f(u), f(v)) = \mu_{(X, \leq)} (u, v) \quad (u, v \in X).$$

*Proof.* This follows, for instance, from 4.48. For, the chains of a given length from  $u$  to  $v$  in  $(X, \leq)$  correspond bijectively to the chains of that length from  $f(u)$  to  $f(v)$  in  $(X', \leq')$ ; we merely apply  $f$  to each element in the chain.  $\square$

**4.58. Example: Möbius Function of a Boolean Poset.** Consider again the poset  $X = (\mathcal{P}([n]), \subseteq)$ . For  $1 \leq i \leq n$ , take  $Y_i = \{0, 1\}$  with the usual ordering, and let  $Y = Y_1 \times \cdots \times Y_n$  be the product poset. There is a bijection  $f$  from  $\mathcal{P}([n])$  to  $\{0, 1\}^n$  that sends a subset  $S$  to the word  $f(S) = w = w_1 w_2 \cdots w_n$  with  $w_i = 1$  for  $i \in S$  and  $w_i = 0$  for  $i \notin S$ . One readily sees that  $f$  is a poset isomorphism, so  $\mu_X(S, T) = \mu_Y(f(S), f(T))$ . Writing  $f(T) = z = z_1 z_2 \cdots z_n$ , 4.55 shows that  $\mu_Y(w, z) = \prod_{i=1}^n \mu_{Y_i}(w_i, z_i)$ . As in 4.44, we see that

$$\mu_{Y_i}(0, 0) = \mu_{Y_i}(1, 1) = 1; \quad \mu_{Y_i}(0, 1) = -1; \quad \mu_{Y_i}(1, 0) = 0.$$

So  $\mu_Y(w, z) = 0$  unless  $w \leq z$ . If  $w \leq z$  and  $z$  has  $k$  more ones than  $w$  does, we see that  $\mu_Y(w, z) = (-1)^k$ . Translating back to subsets via  $f^{-1}$ , this says that  $\mu_X(S, T) = 0$  when  $S \not\subseteq T$ , and  $\mu_X(S, T) = (-1)^{|T \setminus S|}$  when  $S \subseteq T$ .

**4.59. Example: Möbius Function of a Divisibility Poset.** Let  $n$  be a fixed positive integer with prime factorization  $n = p_1^{n_1} \cdots p_k^{n_k}$ , and consider the divisibility poset  $(X, |)$ , where  $X = \{d : d|n\}$ . For  $1 \leq i \leq k$ , let  $Y_i = \{0, 1, \dots, n_i\}$  with the usual ordering, and take  $Y$  to be the product poset  $Y_1 \times \cdots \times Y_k$ . Any  $d \in X$  has prime factorization  $d = p_1^{d_1} \cdots p_k^{d_k}$  for some  $d_k \leq n_k$ . The map  $d \mapsto (d_1, \dots, d_k)$  is readily seen to be a poset isomorphism from  $X$  to  $Y$ . So

$$\mu_X(d, e) = \mu_Y((d_1, \dots, d_k), (e_1, \dots, e_k)) = \prod_{i=1}^k \mu_{Y_i}(d_i, e_i).$$

As in 4.44, we see that  $\mu_{Y_i}(d_i, e_i) = \chi(e_i = d_i) - \chi(e_i = d_i + 1)$ . It follows that  $\mu_X(d, e) = 0$  unless  $e$  is obtained from  $d$  by multiplying by a set of  $s$  *distinct* prime factors chosen from  $\{p_1, \dots, p_k\}$ , in which case  $\mu_X(d, e) = (-1)^s$ . It is now routine to check that whenever  $d|e$ ,  $\mu_X(d, e) = \mu(e/d)$ , where  $\mu$  is the number-theoretic Möbius function.

## Summary

- *Involutions.* An involution is a function  $I : X \rightarrow X$  with  $I \circ I = \text{id}_X$ . The fixed point set of  $I$  is  $\text{Fix}(I) = \{x \in X : I(x) = x\}$ . If  $X$  consists of signed objects,  $I$  is sign-reversing iff  $\text{sgn}(I(x)) = -\text{sgn}(x)$  for all  $x \in X \sim \text{Fix}(I)$ . For a sign-reversing involution  $I$  with domain  $X$ ,

$$\sum_{x \in X} \text{sgn}(x) = \sum_{x \in \text{Fix}(I)} \text{sgn}(x).$$

Involutions provide combinatorial proofs of identities that involve signed terms.

- *Inclusion-Exclusion Formulas.* For arbitrary finite sets  $S_1, \dots, S_n$ ,

$$|S_1 \cup S_2 \cup \dots \cup S_n| = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} |S_{i_1} \cap S_{i_2} \cap \dots \cap S_{i_k}|.$$

If each  $S_i$  is a subset of a finite set  $X$ , then

$$|X \sim (S_1 \cup \dots \cup S_n)| = \sum_{I \subseteq \{1, 2, \dots, n\}} (-1)^{|I|} \left| \bigcap_{i \in I} S_i \right|,$$

where the summand indexed by  $I = \emptyset$  is interpreted as  $|X|$ . In the special case where  $|\bigcap_{i \in I} S_i| = N(k)$  for all  $k$ -element subsets  $I$ , the formula simplifies to

$$|X \sim (S_1 \cup \dots \cup S_n)| = |X| + \sum_{k=1}^n (-1)^k \binom{n}{k} N(k).$$

- *Surjections and Stirling Numbers.* For  $m \geq n \geq 1$ , there are  $\sum_{k=0}^n (-1)^k \binom{n}{k} (n-k)^m$  surjections from an  $m$ -element set onto an  $n$ -element set. A summation formula for the Stirling number of the second kind is

$$S(m, n) = \sum_{k=0}^n (-1)^k \frac{(n-k)^m}{k!(n-k)!}.$$

- *Euler's  $\phi$  Function.* For  $m \geq 1$ ,  $\phi(m)$  is the number of positive integers  $x \leq m$  with  $\gcd(x, m) = 1$ . We have  $\phi(m) = m \prod_{p|m} (1 - p^{-1})$  where the product ranges over all prime divisors  $p$  of  $m$ . For  $m = q^e$  with  $q$  prime,  $\phi(q^e) = q^e - q^{e-1}$ . If  $\gcd(m, n) = 1$ , then  $\phi(mn) = \phi(m)\phi(n)$ . For  $m \geq 1$ ,  $\sum_{d|m} \phi(d) = m$ .
- *Derangements.* A *derangement* of  $S$  is a bijection  $f: S \rightarrow S$  with  $f(x) \neq x$  for all  $x \in S$ . Let  $d_n$  be the number of derangements of an  $n$ -element set. Then  $d_n = n! \sum_{k=0}^n (-1)^k / k!$  is the closest integer to  $n!/e$ . Moreover, the numbers  $d_n$  satisfy the recursions

$$d_n = (n-1)d_{n-1} + (n-1)d_{n-2} \quad (n \geq 2);$$

$$d_n = nd_{n-1} + (-1)^n \quad (n \geq 1).$$

- *Coefficients of Chromatic Polynomials.* For a simple graph  $G$ , the coefficient of  $x^c$  in the chromatic polynomial  $\chi_G(x)$  is  $\sum_{e \geq 0} (-1)^e n(e, c)$ , where  $n(e, c)$  is the number of subgraphs  $H$  of  $G$  such that  $V(H) = V(G)$ ,  $|E(H)| = e$ , and  $H$  has  $c$  connected components.
- *Number-theoretic Möbius Function.* Define  $\mu: \mathbb{N}^+ \rightarrow \{-1, 0, 1\}$  by  $\mu(n) = (-1)^s$  if  $n$  is the product of  $s \geq 0$  distinct primes, and  $\mu(n) = 0$  otherwise. Then  $\sum_{d|m} \mu(d) = \chi(m=1)$ . Given functions  $f$  and  $g$  such that  $f(m) = \sum_{d|m} g(d)$  for  $m \geq 1$ , the classical Möbius inversion formula states that

$$g(m) = \sum_{d|m} f(m/d) \mu(d) = \sum_{d|m} f(d) \mu(m/d) \quad (m \geq 1).$$

It follows that  $\phi(m) = \sum_{d|m} \mu(d) m/d$ .

- *Posets.* A *partial ordering* of  $X$  is a relation  $\leq$  on  $X$  that is reflexive, antisymmetric, and transitive; the pair  $(X, \leq)$  is called a *poset*. A *strict ordering* of  $X$  is a relation  $<$  on  $X$  that is irreflexive and transitive. There is a bijection between partial orders on  $X$  and strict orders on  $X$  defined by removing the diagonal  $\{(x, x) : x \in X\}$ . A *chain of length  $k$*  in a poset  $(X, \leq)$  is a sequence  $(z_0, z_1, \dots, z_k)$  with  $z_0 < z_1 < \dots < z_k$ . Such a chain goes from  $z_0$  to  $z_k$  and has *sign*  $(-1)^k$ .
- *Möbius Functions for Posets.* Given a poset  $(X = \{x_1, \dots, x_n\}, \leq)$ , define  $n \times n$  matrices  $Z$ ,  $N$ , and  $M$  by  $Z_{ij} = \chi(x_i \leq x_j)$ ,  $N_{ij} = \chi(x_i < x_j)$ , and  $M_{ij}$  = the signed sum of all chains in the poset from  $x_i$  to  $x_j$ . Then  $Z = I + N$ ;  $N$  is nilpotent; and  $M$  is the matrix inverse of  $Z$ . We write  $\mu_X(x_i, x_j) = M_{ij}$  and call  $\mu$  the *Möbius function* of the poset  $(X, \leq)$ . Suppose  $f$  and  $g$  are functions with domain  $X$ . The *Möbius inversion formula for posets* states that

$$g(x) = \sum_{y \leq x} f(y) \text{ for all } x \in X \text{ iff } f(x) = \sum_{y \leq x} g(y) \mu_X(y, x) \text{ for all } x \in X.$$

- *Product Posets.* Given posets  $(X_i, \leq_i)$  for  $1 \leq i \leq n$ , the product set  $X = X_1 \times \dots \times X_n$  becomes a poset by defining  $(x_1, \dots, x_n) \leq (y_1, \dots, y_n)$  iff  $x_i \leq_i y_i$  for all  $i$ . The Möbius function for the product poset satisfies

$$\mu_X((x_1, \dots, x_n), (y_1, \dots, y_n)) = \prod_{i=1}^n \mu_{X_i}(x_i, y_i).$$

- *Examples of Möbius Functions.* The poset  $X = \{1, 2, \dots, n\}$  with the usual total ordering has Möbius function

$$\mu_X(i, i) = 1, \quad \mu_X(i, i+1) = -1, \quad \mu_X(i, j) = 0 \text{ for } j \neq i, i+1.$$

The Boolean poset  $(\mathcal{P}(X), \subseteq)$  of subsets of  $\{1, 2, \dots, n\}$  ordered by inclusion has Möbius function

$$\mu(S, T) = (-1)^{|T \setminus S|} \chi(S \subseteq T) \quad (S, T \subseteq [n]).$$

If  $N$  has prime factorization  $p_1^{n_1} \dots p_k^{n_k}$ , then the poset of divisors of  $N$  under the divisibility ordering has Möbius function

$$\mu(d, e) = \begin{cases} (-1)^s & \text{if } e/d \text{ is a product of } s \text{ distinct primes;} \\ 0 & \text{otherwise.} \end{cases}$$

These results follow since the Boolean poset is isomorphic to the product of  $n$  copies of the totally ordered set  $\{0, 1\}$ , whereas the divisibility poset is isomorphic to the product poset  $\{0, 1, \dots, n_1\} \times \dots \times \{0, 1, \dots, n_k\}$ .

## Exercises

**4.60.** Given that  $|S| = 15$ ,  $|T| = 13$ ,  $|U| = 12$ ,  $|S \cap T| = 6$ ,  $|S \cap U| = 3$ ,  $|T \cap U| = 4$ , and  $|S \cap T \cap U| = 1$ , find: (a)  $|S \cup T|$ ; (b)  $|S \cup T \cup U|$ ; (c) the number of objects in exactly one of the sets  $S, T, U$ .

**4.61.** Given that  $S, T, U$  are subsets of  $X$  with  $|X| = 35$ ,  $|S| = 12$ ,  $|T| = 14$ ,  $|U| = 15$ ,  $|S \cap T| = 5 = |S \cap U|$ ,  $|T \cap U| = 6$ , and  $|(S \cup T) \cap U| = 9$ , find: (a)  $|S \cap T \cap U|$ ; (b)  $|X \sim (S \cup T \cup U)|$ ; (c) the number of objects in exactly two of the sets  $S, T, U$ .

**4.62.** List all the derangements in  $D_4$ .

**4.63.** Compute  $d_{10}$  in four ways: (a) by rounding  $10!/e$  to the nearest integer; (b) by using the summation formula 4.23; (c) by using the recursion in 4.24; (d) by using the recursion in 4.25.

**4.64.** Compute  $\phi(n)$ ,  $\mu(n)$ ,  $\tau(n)$ , and  $\sigma(n)$  for the following choices of  $n$ : (a) 6; (b) 11; (c) 28; (d) 60; (e) 1001; (f) 121.

**4.65.** Verify 4.34 by direct calculation for (a)  $m = 24$ ; (b)  $m = 30$ .

**4.66.** Given  $n$  married couples, how many ways can the  $n$  men and  $n$  women be paired up so that no pair consists of a man and his wife?

**4.67.** How many five-card poker hands have at least one card of every suit?

**4.68.** How many five-card poker hands have at least one face card, at least one diamond, and do not contain both a 2 and a 3?

**4.69.** How many ten-digit numbers contain at least one 4, one 5, and one 7?

**4.70.** How many bridge hands are void in clubs and have at least one card of value  $p$  for each prime  $p < 10$ ?

**4.71.** How many surjections  $f : \{1, 2, \dots, m\} \rightarrow \{1, 2, \dots, n\}$  have the property that  $f(x) = 1$  for exactly one  $x \leq m$ ?

**4.72.** (a) What is the chromatic polynomial for the 4-cycle  $C_4$ ? (b) For each coefficient of this chromatic polynomial, draw the vertex-spanning subgraphs of  $C_4$  counted by that coefficient.

**4.73.** For even  $n \geq 2$ , determine the number of integers  $x \leq n$  with  $\gcd(x, n) = 2$ .

**4.74.** For  $k \geq 0$  and  $m \geq 1$ , let  $\sigma_k(m) = \sum_{d|m} d^k$ . (a) Find a formula for  $\sigma_k(m)$  in terms of the prime factorization of  $m$ . (b) Find a formula for  $m^k$  involving  $\sigma_k$  and  $\mu$ .

**4.75.** Use 4.20 to show that  $\phi(mn) = \phi(m)\phi(n)$  iff  $\gcd(m, n) = 1$ .

**4.76.** Explicitly compute how the first involution discussed in 4.6 matches up the 24 objects counted by  $\sum_{k=1}^4 s(4, k)$  into pairs of objects of opposite sign.

**4.77.** Suppose  $w$  has cycles  $(1)$ ,  $(2)$ ,  $(3, 8, 7)$ ,  $(5, 6, 9)$ ,  $(4)$ , and

$$U = \{\{(1)\}, \{(2)\}, \{(4), (5, 6, 9)\}, \{(3, 8, 7)\}\}.$$

Compute  $I(w, U)$ , where  $I$  is the involution defined at the end of 4.6.

**4.78.** Consider the derangement  $w = 436215 \in D_6$ . Find the six derangements in  $D_7$  and the seven derangements in  $D_8$  that can be built from  $w$  by the construction in the proof of 4.24.

**4.79.** Use the recursion for derangements in 4.25 to give a proof by induction of the summation formula for derangements in 4.23.

**4.80.** Give the details of the proof of 4.39.

**4.81.** Show that if  $G$  is a simple graph with  $c$  connected components, then the chromatic polynomial  $\chi_G(x)$  must be divisible by  $x^c$ .

**4.82.** (a) Give an algebraic proof that  $\sum_{k=0}^n \binom{n}{k} 2^k (-1)^{n-k} = 1$  for  $n \geq 0$ . (b) Prove the identity in (a) using an involution.

**4.83.** For integers  $a \geq b > 0$ , evaluate  $\sum_{k=0}^n \binom{n}{k} a^{n-k} (-b)^k$  by using an involution.

**4.84.** For integers  $0 \leq a \leq b \leq c$ , evaluate  $\sum_{k=a}^b (-1)^k \binom{c}{k}$  by using an involution.

**4.85.** Let  $S \subseteq T$  be given finite sets. (a) Use an involution to prove  $\sum_{U: S \subseteq U \subseteq T} (-1)^{|T \sim U|} = \chi(S = T)$  (cf. 4.45). (b) In a similar manner, evaluate  $\sum_{U: S \subseteq U \subseteq T} (-1)^{|U \sim S|}$ .

**4.86.** Let  $d, e \in \mathbb{N}^+$  with  $d|e$ . Use an involution to prove  $\sum_{k: d|k|e} \mu(e/k) = \chi(d = e)$ . Interpret this result in terms of the Möbius function of a poset.

**4.87.** Count the  $n \times n$  matrices  $A$  with entries in  $\{0, 1, 2\}$  such that: (a) no row of  $A$  contains all zeroes; (b) every column of  $A$  contains at least one zero; (c) there is no index  $j$  with  $A(i, j) > 0$  and  $A(j, i) > 0$  for all  $i$ .

**4.88.** An *arrowless vertex* in a simple digraph  $D$  is a vertex with indegree and outdegree zero. How many simple digraphs with vertex set  $\{1, 2, \dots, n\}$  have no arrowless vertices?

**4.89.** An *isolated vertex* in a simple digraph  $D$  is a vertex  $v$  such that there is no edge  $(u, v)$  or  $(v, u)$  in  $D$  with  $u \neq v$ . How many simple digraphs with vertex set  $\{1, 2, \dots, n\}$  have no isolated vertices?

**4.90.** How many simple graphs with vertex set  $\{1, 2, \dots, n\}$  have no isolated vertices?

**4.91.** Use 4.11 to compute the chromatic polynomial of the paw graph (see 3.124).

**4.92.** (a) How many anagrams in  $\mathcal{R}(1^3 2^3 \dots n^3)$  never have three equal letters in a row? (b) How many anagrams in  $\mathcal{R}(1^k 2^k \dots n^k)$  never have  $k$  equal letters in a row?

**4.93.** (a) Count the permutations  $w$  of  $\{1, 2, \dots, n\}$  such that  $w_{i+1} \neq w_i + 1$  for all  $i < n$ . (b) Express your answer to (a) in terms of the derangement numbers  $d_k$ .

**4.94.** Given sequences  $0 \leq a_1 \leq a_2 \leq \dots \leq a_k \leq A$  and  $0 \leq b_1 \leq b_2 \leq \dots \leq b_k \leq B$ , use inclusion-exclusion to derive a formula for the number of lattice paths from  $(0, 0)$  to  $(A, B)$  that avoid all of the points  $(a_i, b_i)$  for  $1 \leq i \leq k$ .

**4.95. Recursion for Möbius functions.** (a) Show that the Möbius function of a poset  $(X, \leq)$  can be computed recursively via  $\mu(x, z) = -\sum_{y: x \leq y < z} \mu(x, y)$  for  $x < z$ , with initial conditions  $\mu(x, x) = 1$  and  $\mu(x, z) = 0$  whenever  $x \not\leq z$ . (b) Show that the Möbius function also satisfies the recursion  $\mu(x, z) = -\sum_{y: x < y \leq z} \mu(y, z)$  for  $x < z$ .

**4.96. Poset Associated to a DAG.** Suppose  $G = (X, R)$  is a DAG. Prove that there exists a unique smallest irreflexive, transitive relation  $<$  that contains  $R$ . The corresponding poset  $(X, \leq)$  is called the *poset associated to the DAG*  $G$ .

**4.97.** Let  $(X, \leq)$  be the poset associated to the DAG

$$(\{a, b, c, d, e\}, \{(a, b), (b, e), (a, c), (c, e), (a, d), (d, e)\}).$$

Compute the Möbius function  $\mu_X$  in two ways, by: (a) inverting the matrix  $Z$ ; (b) enumerating signed chains in  $(X, \leq)$ .



**4.98.** Let  $(X, \leq)$  be the poset associated to the DAG

$$(\{a, b, c, d, e, f\}, \{(a, b), (a, c), (b, d), (b, e), (c, d), (d, f), (e, f)\}).$$

Compute the Möbius function  $\mu_X$  in two ways, by: (a) inverting the matrix  $Z$ ; (b) enumerating signed chains in  $(X, \leq)$ .

**4.99.** A *subposet* of a poset  $(X, \leq)$  is a poset  $(Y, \leq')$ , where  $Y$  is a subset of  $X$ , and for  $a, b \in Y$ ,  $a \leq' b$  iff  $a \leq b$ . An *interval* in  $X$  is a subposet of the form  $[x, z] = \{y \in X : x \leq y \leq z\}$ . Show that for all  $a, b, c, d \in X$ , if the intervals  $[a, b]$  and  $[c, d]$  are isomorphic posets, then  $\mu_X(a, b) = \mu_X(c, d)$ .

**4.100.** Assume that  $X_1$  and  $X_2$  are finite disjoint sets. The *disjoint union* of the posets  $(X_1, \leq_1)$  and  $(X_2, \leq_2)$  is  $(X, \leq)$  where  $X = X_1 \cup X_2$  and for  $a, b \in X$ ,  $a \leq b$  iff  $a, b \in X_1$  and  $a \leq_1 b$ , or  $a, b \in X_2$  and  $a \leq_2 b$ . Determine  $\mu_X$  in terms of  $\mu_{X_1}$  and  $\mu_{X_2}$ .

**4.101.** Given a poset  $(X, \leq)$ , define a new poset  $(Y, \leq')$  by setting  $Y = X \cup \{0\}$  (where 0 is a new symbol not in  $X$ ), and letting  $\leq'$  be the extension of  $\leq$  such that  $0 \leq' y$  for all  $y \in Y$ . Informally,  $(Y, \leq')$  is obtained from  $(X, \leq)$  by adjoining a new least element. Determine  $\mu_Y$  in terms of  $\mu_X$ .

**4.102.** Given posets  $(X_1, \leq_1)$  and  $(X_2, \leq_2)$  where  $X_1$  and  $X_2$  are finite disjoint sets, define a new poset  $(X, \leq)$  by setting  $X = X_1 \cup X_2$  and, for  $a, b \in X$ ,  $a \leq b$  iff  $a, b \in X_i$  and  $a \leq_i b$  ( $i = 1, 2$ ), or  $a \in X_1$  and  $b \in X_2$ . Informally,  $(X, \leq)$  is obtained from  $X_1$  and  $X_2$  by making everything in  $X_1$  less than everything in  $X_2$ . Determine  $\mu_X$  in terms of  $\mu_{X_1}$  and  $\mu_{X_2}$ .

**4.103.** Let  $S_1, \dots, S_n$  be any events in a sample space  $X$  with probability measure  $P$ . State and prove an analogue of 4.7 that can be used to compute  $P(S_1 \cup \dots \cup S_n)$ .

**4.104.** Let  $S_1, \dots, S_n$  be independent events in a sample space  $X$  (see 1.84). Prove that for  $1 \leq i \leq n$ , the events  $S_1, S_2, \dots, X \sim S_i, \dots, S_n$  are independent.

**4.105.** Let  $S_1, \dots, S_n$  be independent events in a sample space  $X$ , with  $P(S_i) = p_i$  for each  $i$ . Find the probability that none of the events  $S_i$  occurs: (a) using inclusion-exclusion and the generalized distributive law; (b) using 4.104.

**4.106.** Use an involution to prove that for all  $i, n \in \mathbb{N}$ ,  $\sum_{k=0}^i (-1)^k \binom{n}{k} \binom{n-k}{i-k} = \chi(i=0)$ .

**4.107.** Use an involution to prove that for  $0 \leq k \leq n$ ,  $\sum_{i=k}^n (-1)^{i-k} \binom{n}{i} \binom{i}{k} 2^{n-i} = \binom{n}{k}$ .

**4.108.** Prove that for all  $n, j > 0$ ,  $n^j = \sum_{k=0}^j (-1)^{j-k} k! S(j, k) \binom{n+k-1}{k}$ .

**4.109.** For  $n > 0$ , evaluate  $\sum_{k=0}^{n-1} (-1)^k \binom{n}{k} (n-k)^n$ .

**4.110.** Use an involution to prove the following identity satisfied by Catalan numbers:  $C_n = \sum_{1 \leq k \leq (n+1)/2} (-1)^{k-1} C_{n-k} \binom{n+1-k}{k}$ .

**4.111.** Let  $A$  be an  $n \times n$  matrix with  $A(i, j) = \binom{i-1}{j-1}$  for  $1 \leq i, j \leq n$ . (a) Look at small examples to guess a formula for  $A^{-1}(i, j)$ . (b) Prove your guess using an involution.

**4.112.** How many bijections  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  are such that the functional digraph of  $f$  contains no cycle of length  $k$ ?

**4.113.** How many anagrams in  $\mathcal{R}(a^3 b^3 c^3 d^3)$  never have two consecutive equal letters?

**4.114.** Prove or disprove: for every integer  $y \geq 1$ , there exist only finitely many integers  $x \geq 1$  with  $\phi(x) = y$ .

**4.115.** How many compositions of  $n$  have  $k$  parts each of size at most  $m$ ?

**4.116.** Call a function  $f : X \rightarrow Y$  *doubly surjective* iff for all  $y \in Y$ , there exist at least two  $x \in X$  with  $f(x) = y$ . Count the number of doubly surjective functions from an  $m$ -element set to an  $n$ -element set, where  $m \geq 2n$ . What is the answer when  $m = 11$  and  $n = 4$ ?

**4.117.** (a) Let  $S_1, \dots, S_n$  be subsets of a finite set  $X$ . Prove that the number of elements of  $X$  that lie in exactly  $k$  of the sets  $S_i$  is

$$\sum_{i=0}^{n-k} (-1)^i \binom{k+i}{i} \sum_{1 \leq j_1 < j_2 < \dots < j_{k+i} \leq n} |S_{j_1} \cap \dots \cap S_{j_{k+i}}|.$$

(b) Find and prove a similar formula for the number of elements of  $X$  that lie in at least  $k$  of the sets  $S_i$ .

**4.118.** For  $0 \leq k \leq n$ , let  $d_{n,k}$  be the number of permutations of  $n$  objects that have exactly  $k$  fixed points. (a) Use 4.117 to find a formula for  $d_{n,k}$ . (b) Give algebraic and combinatorial proofs that  $d_{n,k} = \binom{n}{k} d_{n-k}$ .

**4.119.** How many integers between 1 and 2311 are divisible by exactly two of the primes in  $\{2, 3, 5, 7\}$ ? (Use 4.117.)

**4.120.** Let  $(F_n)$  be the Fibonacci sequence ( $F_0 = 0$ ,  $F_1 = 1$ ,  $F_n = F_{n-1} + F_{n-2}$  for  $n \geq 2$ ). Find a formula for  $\sum_{k=0}^n (-1)^k F_k$  and prove it, either algebraically or using an involution.

**4.121.** Find and prove a formula for  $\sum_{k=0}^n (-1)^k F_k F_{n-k}$ .

**4.122.** For each integer  $x \geq 1$ , evaluate  $\sum_{k=1}^x \mu(k) \lfloor x/k \rfloor$ .

**4.123.** For  $n > 0$ , evaluate  $\sum_{k=1}^n (-1)^k \text{Surj}(n, k)$ .

**4.124.** For  $n > 0$ , evaluate  $\sum_{k=1}^{n-1} (-1)^k (k-1)! S(n, k)$ .

**4.125.** Consider an  $n \times n$  lower-triangular matrix  $A$  such that  $A(n, k)$  is the number of Dyck paths ending with exactly  $k$  east steps, for  $1 \leq k \leq n$ . Find a combinatorial description of  $A^{-1}$ , and prove that this is the inverse of  $A$  using an involution.

**4.126. Garsia-Milne Involution Principle.** Suppose  $I$  and  $J$  are involutions defined on finite signed sets  $X$  and  $Y$ , respectively. Suppose  $f : X \rightarrow Y$  is a *sign-preserving* bijection, i.e.,  $\text{sgn}(f(x)) = \text{sgn}(x)$  for all  $x \in X$ . Suppose also that every object in  $\text{Fix}(I)$  and  $\text{Fix}(J)$  has positive sign. Construct an explicit bijection  $g : \text{Fix}(I) \rightarrow \text{Fix}(J)$ .

**4.127. Bijective Subtraction.** Suppose  $A$ ,  $B$ , and  $C$  are finite, pairwise disjoint sets and  $f : A \cup B \rightarrow A \cup C$  is a given bijection. Construct an explicit bijection  $g : B \rightarrow C$ .

**4.128. Bijective Division by Two.** Suppose  $A$  and  $B$  are finite sets. Given a bijection  $f : \{0, 1\} \times A \rightarrow \{0, 1\} \times B$ , can you use  $f$  to construct an explicit bijection  $g : A \rightarrow B$ ?

**4.129.** In §4.1 we proved combinatorially that  $\sum_k s(i, k) S(k, j) = \chi(i = j)$ . Can you find a combinatorial proof that  $\sum_k S(i, k) s(k, j) = \chi(i = j)$ ? (Compare with 2.77(d).)

**4.130.** Find a bijective proof of the derangement recursion  $d_n = n d_{n-1} + (-1)^n$ .

**4.131.** Let  $X_n$  be the set of set partitions of  $\{1, 2, \dots, n\}$ . Define the *refinement ordering* on  $X_n$  by setting, for  $P, Q \in X_n$ ,  $P \preceq Q$  iff every block  $S \in P$  is contained in some block  $T \in Q$ . (a) Show that  $(X_n, \preceq)$  is a poset. (b) Compute the Möbius function of this poset for  $1 \leq n \leq 4$ . (c) Show that any interval  $[P, Q]$  in  $X_n$  (see 4.99) is isomorphic to a poset  $(X_k, \preceq)$  for some  $k$ . (d) Compute  $\mu_{X_n}$  for all  $n$ .

---

## Notes

A thorough treatment of posets from the combinatorial viewpoint appears in Chapter 3 of Stanley [127]. See Rota [118] for one of the seminal papers on Möbius inversion in combinatorics. A classic text on posets is the book by Birkhoff [12]. The Garsia-Milne involution principle in 4.126 was introduced in [49, 50]. For applications and extensions of this principle, the reader may consult the following sources [57, 73, 87, 88, 108, 109, 140]. An application of bijective subtraction (see 4.127) is presented in Loehr [85].

---

## Ranking and Unranking

---

This chapter studies the notions of *ranking* and *unranking* from a bijective viewpoint. Intuitively, our goal is to find algorithms that implement bijective maps between an  $n$ -element set of combinatorial objects and the set of integers  $\underline{n} = \{0, 1, 2, \dots, n-1\}$ . These algorithms will allow us to solve a variety of combinatorial problems. We begin the chapter by introducing some of these problems. Then we discuss bijective versions of the sum and product rules. These new rules provide a mechanical method for translating the counting arguments in earlier chapters into explicit bijections. Recursions derived using the sum and product rules can be treated in the same way; the resulting bijections are typically specified by recursive algorithms.

---

### 5.1 Ranking, Unranking, and Related Problems

Suppose  $S$  is a finite set of objects. We will study five fundamental combinatorial problems involving the set  $S$ : counting, listing, ranking, unranking, and random selection.

1. *Counting.* The counting problem asks us to compute the number of elements in the finite set  $S$ . We have already discussed tools to solve the counting problem in Chapter 1.
2. *Listing.* The listing problem asks us to list all the elements in the set  $S$  exactly once. There are many possible lists for a given set  $S$ ; usually we produce lists that present the objects of  $S$  in a special order. Examples of such orderings are lexicographic orderings and Gray code orderings.
3. *Ranking.* Suppose we have specified a particular ordering for listing the elements of  $S$ . Given an object  $x \in S$ , the ranking problem asks us to calculate the position (or *rank*) of  $x$  on the list without actually listing all the objects preceding (or following)  $x$  on the list. It is often convenient to number the positions on the list starting with position 0.
4. *Unranking.* Suppose we have specified a particular ordering for listing the elements of  $S$ . Given an integer  $m$  with  $0 \leq m < |S|$ , the unranking problem asks us to calculate the object  $x$  in  $S$  that occupies position  $m$  on the list, without actually generating the whole list.
5. *Random Selection.* The random selection problem asks us to devise a way to choose a random element of  $S$ , where “random” means that each element of  $S$  is equally likely to be chosen. We assume that we are given a device that will produce random real numbers in the interval  $[0, 1]$ . Alternatively, we can assume that we have a device that, given a positive integer  $n$ , randomly picks an integer in the set  $\underline{n} = \{0, 1, \dots, n-1\}$ .

If we can solve the listing problem for  $S$ , then we can (in principle) solve the counting problem, the ranking problem, and the unranking problem by simply writing down the whole

list. However, we usually desire more efficient methods for counting  $S$ , ranking objects, and unranking integers. Also, this method is impractical if we are studying not just one finite set  $S$  but a whole infinite family of such sets.

Similarly, if we can solve the counting problem and unranking problem for  $S$ , then we can solve the random selection problem as follows. Use the random number generator described above, with  $n = |S|$ , to get a random integer between 0 and  $|S| - 1$ . Unrank this integer to get the random object in  $S$ . Note that if  $|S|$  is very large, it may be difficult to implement a random number generator that uniformly chooses integers in the desired range. Thus, even if we can count  $S$  and unrank elements of  $S$ , it is valuable to find other ways to solve the random selection problem that avoid the random selection of integers in a huge range.

In bijective combinatorics, we try to solve the five enumeration problems posed above by constructing explicit *bijections*. For example, the ranking problem amounts to constructing a bijection  $r : S \rightarrow \underline{n}$  with specified properties. The unranking problem amounts to constructing the inverse bijection  $u : \underline{n} \rightarrow S$ . The list of elements of  $S$  determined by the unranking map  $u$  is the sequence  $(u(0), u(1), \dots, u(n-1))$ .

Here and below, “constructing” a map  $h : A \rightarrow B$  means giving an algorithm that takes as input an element  $a \in A$  and produces as output the corresponding element  $h(a) \in B$ . We must also prove that the proposed algorithm is indeed a bijection, and we would like to have algorithms that are as efficient as possible. Note that knowing an algorithm to compute a bijection  $h : A \rightarrow B$  is not the same as knowing an algorithm to compute the inverse bijection  $h^{-1} : B \rightarrow A$ . We say that we have solved the counting problem for  $S$  *bijectively* if we have algorithms to compute both a bijection  $u : \underline{n} \rightarrow S$  and its inverse  $r : S \rightarrow \underline{n}$ . Observe that the bijective counting problem is harder than the original enumerative counting problem: if  $S$  is complicated, we may be able to determine that  $|S| = n$  without constructing any explicit bijections between  $S$  and  $\underline{n}$ . By definition, saying that  $|S| = n$  means that such bijections *exist*. But knowing this abstract existence statement is much weaker than actually giving constructions and algorithms that implement particular bijections.

As noted above, each bijection  $u : \underline{n} \rightarrow S$  provides one solution to the listing problem for  $S$ . If we are asked to list elements of  $S$  in a particular order, then we must construct an appropriate bijection  $u$  such that the list  $(u(0), u(1), \dots, u(n-1))$  contains the objects in  $S$  in the desired order. It is sometimes desirable to have an auxiliary *successor* algorithm that, given an object  $u(i)$  in the list, outputs the next object  $u(i+1)$  without ever explicitly computing  $i$ . We can then list the objects in  $S$  by starting with  $u(0)$  and repeatedly invoking the successor algorithm. We will consider the construction of successor algorithms at the end of this chapter.

## 5.2 Bijective Sum Rule

We begin our study of ranking and unranking by revisiting the fundamental counting rules from Chapter 1. Our first rule lets us assemble ranking (resp. unranking) maps for two disjoint finite sets to obtain a ranking (resp. unranking) map for the union of these sets. Throughout this chapter, the notation  $\underline{n}$  will be used to denote the set  $\{0, 1, \dots, n-1\}$ .

**5.1. Bijective Sum Rule for Two Sets.** Let  $S$  and  $T$  be disjoint finite sets.

(a) Given bijections  $f : S \rightarrow \underline{n}$  and  $g : T \rightarrow \underline{m}$ , there is a canonical bijection  $f + g : S \cup T \rightarrow \underline{n + m}$  defined by

$$(f + g)(x) = \begin{cases} f(x) & \text{for } x \in S; \\ g(x) + n & \text{for } x \in T. \end{cases}$$

(b) Given bijections  $f' : \underline{n} \rightarrow S$  and  $g' : \underline{m} \rightarrow T$ , there is a canonical bijection  $f' + g' : \underline{n + m} \rightarrow S \cup T$  defined by

$$(f' + g')(k) = \begin{cases} f'(k) & \text{for } 0 \leq k < n \\ g'(k - n) & \text{for } n \leq k < n + m \end{cases}$$

(c) If  $f' = f^{-1}$  and  $g' = g^{-1}$ , then  $f' + g' = (f + g)^{-1}$ .

We leave the detailed verification of this rule as an exercise. The disjointness of  $S$  and  $T$  is critical when showing that  $f + g$  is a well-defined function and that  $f' + g'$  is injective. Observe that the *order* in which we combine the bijections makes a difference: the bijection  $f + g : S \cup T \rightarrow \underline{n + m}$  is not the same as the bijection  $g + f : S \cup T \rightarrow \underline{n + m}$ . Intuitively, the ranking bijection  $f + g$  assigns earlier ranks to elements of  $S$  (using  $f$  to determine these ranks) and assigns later ranks to elements of  $T$  (using  $g$ );  $g + f$  does the opposite. Similarly, the unranking map  $f' + g'$  generates a list in which elements of  $S$  occur first, followed by elements of  $T$ ;  $g' + f'$  lists elements of  $T$  first, then  $S$ .

Iterating the bijective sum rule for two sets leads to the following general version of this rule.

**5.2. Bijective Sum Rule for  $k$  Sets.** Suppose  $(S_1, \dots, S_k)$  is an ordered list of pairwise disjoint finite sets. Let  $S = \bigcup_{i=1}^k S_i$  be the union of these sets. Let  $n_i = |S_i|$  and  $n = |S| = n_1 + n_2 + \dots + n_k$ .

(a) Given bijections  $f_i : S_i \rightarrow \underline{n_i}$  for  $1 \leq i \leq k$ , there is a canonical bijection  $f = \sum_{i=1}^k f_i : S \rightarrow \underline{n}$  defined by  $f(x) = f_i(x) + \sum_{j < i} n_j$  for  $x \in S_i$ .

(b) Given bijections  $f'_i : \underline{n_i} \rightarrow S_i$  for  $1 \leq i \leq k$ , there is a canonical bijection  $f' = \sum_{i=1}^k f'_i : \underline{n} \rightarrow S$  specified as follows. For each  $x \in \underline{n}$ , there exists a unique  $i$  ( $1 \leq i \leq k$ ) such that  $n_1 + \dots + n_{i-1} \leq x < n_1 + \dots + n_i$ . Define  $f'(x) = f'_i(x - [n_1 + \dots + n_{i-1}]) \in S_i \subseteq S$ .

(c) If  $f'_i = f_i^{-1}$  for each  $i$ , then  $f' = f^{-1}$ .

As before, we leave the formal proof of the bijective sum rule to the reader. (One can give a direct proof that the maps in question are bijections, or use an induction argument involving the bijective sum rule for two sets to show that  $\sum_{i=1}^k f_i = \sum_{i=1}^{k-1} f_i + f_k$ .) Intuitively,  $f_1 + \dots + f_k$  is the ranking map that assigns elements of  $S_1$  to positions 0 through  $n_1 - 1$  using  $f_1$ , assigns elements of  $S_2$  to positions  $n_1$  through  $n_1 + n_2 - 1$  using  $f_2$ , etc. The unranking map  $f'_1 + \dots + f'_k$  generates a listing of  $S$  in which objects in  $S_1$  occur first, then objects in  $S_2$ , and so on.

## 5.3 Bijective Product Rule

Now we consider bijective versions of the product rule, which generalize the familiar “base- $b$  expansions” of natural numbers. Before introducing the bijective product rule, we recall the following theorem concerning integer division with remainder.

**5.3. Theorem: Integer Division.** Suppose  $a$  is any integer and  $b$  is a positive integer. There exist unique integers  $q$  and  $r$  such that

$$a = bq + r \text{ and } 0 \leq r < b.$$

Furthermore, there is an algorithm to compute  $q$  and  $r$  given  $a$  and  $b$ . The integers  $q$  and  $r$  are called the *quotient* and *remainder* when  $a$  is divided by  $b$ .

*Proof.* First assume  $a \geq 0$ . Consider the following algorithm. Define  $a_0 = a$  and  $i = 0$ , and then loop as follows. If  $a_i \geq b$ , define  $a_{i+1} = a_i - b \geq 0$ , then replace  $i$  by  $i + 1$ , and continue to loop. Otherwise, terminate with the answer  $q = i$  and  $r = a_i$ . Now, this algorithm must terminate, since otherwise we would have an infinite strictly decreasing sequence  $a_0 > a_1 > a_2 > \cdots$  of elements of  $\mathbb{N}$ , which is impossible. An induction on  $i$  shows that  $a = bi + a_i$  for each  $i$  such that  $a_i$  is defined. Therefore, when the algorithm terminates, we will indeed have  $a = bq + r$  and  $0 \leq r < b$ .

If  $a < 0$ , use the preceding algorithm to compute  $q_1, r_1 \in \mathbb{Z}$  with  $|a| = bq_1 + r_1$  and  $0 \leq r_1 < b$ . Then  $a = -bq_1 - r_1$ . If  $r_1 = 0$ , set  $q = -q_1$  and  $r = 0$ ; otherwise, set  $q = -1 - q_1$  and  $r = b - r_1$ . One readily checks that  $a = bq + r$  and  $0 \leq r < b$ . This completes the algorithmic proof of the existence of  $q$  and  $r$ .

For uniqueness of  $q$  and  $r$ , suppose we have  $a = bq + r = bq' + r'$  where  $q, r, q', r' \in \mathbb{Z}$  and  $0 \leq r, r' < b$ . Rearranging the given equations, we see that  $b(q - q') = r' - r$ . The right side is an integer strictly between  $-b$  and  $b$ , whereas the left side is an integer multiple of  $b$ . The only such multiple of  $b$  is zero, so  $q = q'$  and  $r = r'$ .  $\square$

**5.4. Remark.** The division algorithm used in the preceding proof is quite inefficient. In practice, one divides  $a$  by  $b$  using the “long division” algorithm (see 5.90).

**5.5. Theorem: Base- $b$  Expansions.** Let  $b > 1$  be a fixed positive integer. For every integer  $a \geq 0$ , there exists a unique sequence  $(d_0, d_1, d_2, \dots)$  of integers satisfying the following properties:  $0 \leq d_i < b$  for all  $i$ ; all but finitely many  $d_i$ 's are zero; and

$$a = d_0 + d_1b + d_2b^2 + \cdots + d_ib^i + \cdots = \sum_{i \geq 0} d_ib^i.$$

We call  $(d_0, d_1, \dots)$  the *base- $b$  expansion* of  $a$ , which can be written more concisely as  $[a]_b = \cdots d_3d_2d_1d_0$ .

We only sketch the proof. The idea is to divide  $a$  by  $b$  repeatedly. The first remainder  $r$  is the last “digit”  $d_0$ ; expanding the first quotient  $q$  in the same manner yields the remaining digits  $\cdots d_3d_2d_1$ . Uniqueness follows by induction, using the uniqueness assertion in 5.3. We will give another proof of this result later, by using the bijective product rule to rank and unrank words in the product set  $\mathbf{b}^k$ .

We can generalize the preceding discussion by allowing the base  $b$  to change at each step. This idea leads to the general version of the bijective product rule, presented below. First, we consider the simpler version of this rule involving two sets.

**5.6. Theorem: Bijective Product Rule for Two Sets.** Suppose  $s$  and  $t$  are positive integers, and  $n = st$ . The map  $p = p_{s,t} : \mathbf{s} \times \mathbf{t} \rightarrow \mathbf{n}$  given by  $p(i, j) = it + j$  is a bijection. To compute the inverse map  $p'(u)$  (where  $u \in \mathbf{n}$ ), use division to write  $u = qt + r$  where  $0 \leq r < t$ , and set  $p'(u) = (q, r)$ .

*Proof.* First we check that  $p$  does map  $\mathbf{s} \times \mathbf{t}$  into  $\mathbf{n}$ . Suppose  $0 \leq i < s$  and  $0 \leq j < t$  are integers. Then  $0 \leq it + j$ . Furthermore, since  $i \leq s - 1$ , we have  $it + j \leq (s - 1)t + j < (s - 1)t + t = st$ . Thus,  $p(i, j) \in \mathbf{st} = \mathbf{n}$ . Next we check that  $p'$  does map  $\mathbf{n}$  into  $\mathbf{s} \times \mathbf{t}$ . Given  $u \in \mathbf{n}$ , the integers  $q$  and  $r$  in the definition of  $p'$  are well defined, by the existence and uniqueness assertions in 5.3. The condition on the remainder in 5.3 assures us that  $r \in \mathbf{t}$ . Since  $q < 0$  implies  $u < 0$ , while  $q \geq s$  implies  $u \geq st$ , and we are assuming that  $0 \leq u < n = st$ , we conclude that  $0 \leq q < s$ . Thus,  $q \in \mathbf{s}$ , so  $p'$  does map into the set  $\mathbf{s} \times \mathbf{t}$ . It is now routine to check that the maps  $p$  and  $p'$  are indeed two-sided inverses of each other, so both maps are bijections.  $\square$

**5.7. Remark.** Note that the bijection  $p_{s,t}$  just constructed depends on the order of  $s$  and  $t$ . More explicitly,  $p_{s,t} : \underline{s} \times \underline{t} \rightarrow \underline{n}$  sends  $(i, j)$  to  $it + j$ , whereas  $p_{t,s} : \underline{t} \times \underline{s} \rightarrow \underline{n}$  sends  $(j, i)$  to  $js + i$ . Let  $F : \underline{s} \times \underline{t} \rightarrow \underline{t} \times \underline{s}$  be the bijection given by  $F((i, j)) = (j, i)$ . The preceding formulas show that  $p_{t,s} \circ F \neq p_{s,t}$  for  $s \neq t$ , although both functions are bijections from  $\underline{s} \times \underline{t}$  to  $\underline{st}$ .

**5.8. Remark.** The bijections in the product rule can also be built automatically using the bijective sum rule. In the sum rule, take  $S_i = \{i\} \times \underline{t}$  for  $0 \leq i < s$ , and let  $f_i : S_i \rightarrow \underline{t}$  be the bijection defined by  $f_i(i, j) = j$  for all  $j \in \underline{t}$ . The set  $\underline{s} \times \underline{t}$  is the disjoint union of the  $S_i$ 's, so the bijective sum rule furnishes a bijection  $f = \sum_{i=0}^{s-1} f_i : \underline{s} \times \underline{t} \rightarrow \underline{st}$ . The formula for  $f$  gives  $f(i, j) = f_i(i, j) + \sum_{0 \leq k < i} |S_k| = j + it = p_{s,t}(i, j)$ . Similarly,  $p_{t,s} \circ F$  can be obtained by adding up the bijections  $\underline{s} \times \{j\} \rightarrow \underline{s}$  sending  $(i, j)$  to  $i$ . Since the bijective sum rule guarantees the invertibility of  $f$ , we see that the division algorithm 5.3 can be deduced as a consequence of 5.2.

**5.9. Theorem.** Suppose  $f : A \rightarrow C$  and  $g : B \rightarrow D$  are bijections. Then there is a canonical bijection  $f \times g : A \times B \rightarrow C \times D$  given by

$$(f \times g)(a, b) = (f(a), g(b)) \quad (a \in A, b \in B).$$

*Proof.* First,  $f \times g$  is a function mapping  $A \times B$  into  $C \times D$ . Also  $f^{-1} \times g^{-1}$  (which sends  $(c, d)$  to  $(f^{-1}(c), g^{-1}(d))$ ) is a function from  $C \times D$  to  $A \times B$ . One checks immediately that  $f^{-1} \times g^{-1}$  is the two-sided inverse of  $f \times g$ , so both maps are bijections.  $\square$

This result extends immediately to Cartesian products of finitely many sets.

**5.10. Theorem.** Suppose  $f_i : S_i \rightarrow T_i$  are bijections, for  $1 \leq i \leq k$ . Then the map

$$f = f_1 \times f_2 \times \cdots \times f_k : S_1 \times \cdots \times S_k \rightarrow T_1 \times \cdots \times T_k,$$

given by  $f(s_1, s_2, \dots, s_k) = (f_1(s_1), f_2(s_2), \dots, f_k(s_k))$ , is a bijection with inverse  $f_1^{-1} \times \cdots \times f_k^{-1}$ .

**5.11. Theorem: Bijective Product Rule for  $k$  Sets.** Suppose  $n_1, \dots, n_k$  are given positive integers, and  $n = n_1 n_2 \cdots n_k$ . There are canonical bijections

$$p = p_{n_1, \dots, n_k} : \underline{n_1} \times \underline{n_2} \times \cdots \times \underline{n_k} \rightarrow \underline{n}, \quad p' = p'_{n_1, \dots, n_k} : \underline{n} \rightarrow \underline{n_1} \times \underline{n_2} \times \cdots \times \underline{n_k}.$$

The map  $p$  is defined by

$$p(c_1, c_2, \dots, c_k) = \sum_{i=1}^k c_i \prod_{j=i+1}^k n_j.$$

The inverse map  $p'$  is computed via the following algorithm. Given an input  $m$  with  $0 \leq m < n$ , divide  $m$  by  $n_k$  to get a quotient  $q_k$  and remainder  $r_k$ . Next, divide  $q_k$  by  $n_{k-1}$  to get a quotient  $q_{k-1}$  and remainder  $r_{k-1}$ . Continue in this way, dividing  $q_i$  by  $n_{i-1}$  to get a quotient  $q_{i-1}$  and remainder  $r_{i-1}$ , for  $1 < i \leq k$ . After completing all these divisions, set  $p'(m) = (r_1, r_2, \dots, r_k)$ .

*Proof.* We use induction on  $k$  to show that  $p$  is a bijection. When  $k = 1$ ,  $p$  is an identity map. When  $k = 2$ , the result follows from the bijective product rule for two sets. Finally, assume  $k > 2$  and the result is already known for products of  $k - 1$  sets. Writing  $n' = n_2 n_3 \cdots n_k$ , observe that

$$\begin{aligned} p_{n_1, \dots, n_k}(c_1, \dots, c_k) &= c_1(n_2 \cdots n_k) + \sum_{i=2}^k c_i \prod_{j=i+1}^k n_j \\ &= c_1 n' + p_{n_2, \dots, n_k}(c_2, \dots, c_k) \\ &= p_{n_1, n'}(c_1, p_{n_2, \dots, n_k}(c_2, \dots, c_k)). \end{aligned}$$



This means that  $p_{n_1, \dots, n_k}$  is the composition of the two bijections

$$\text{id}_{\underline{n}_1} \times p_{n_2, \dots, n_k} : \underline{n}_1 \times (\underline{n}_2 \times \dots \times \underline{n}_k) \rightarrow \underline{n}_1 \times \underline{n}' \text{ and } p_{n_1, n'} : \underline{n}_1 \times \underline{n}' \rightarrow \underline{n}.$$

To show that  $p'$  is the inverse of  $p$ , it is sufficient to verify that  $p'(p(c_1, \dots, c_k)) = (c_1, \dots, c_k)$ , since we already know that  $|\underline{n}_1| \times \dots \times |\underline{n}_k| = |\underline{n}|$ . Again we use induction. Note that

$$p_{n_1, \dots, n_k}(c_1, \dots, c_k) = \sum_{i=1}^k c_i \prod_{i < j \leq k} n_j = c_k + n_k \sum_{i=1}^{k-1} c_i \prod_{i < j < k} n_j.$$

This shows that the quotient and remainder when we divide  $p(c_1, \dots, c_k)$  by  $n_k$  are

$$q_k = \sum_{i=1}^{k-1} c_i \prod_{i < j \leq k-1} n_j = p_{n_1, \dots, n_{k-1}}(c_1, \dots, c_{k-1})$$

and  $r_k = c_k$ , respectively. So the first division step successfully recovers  $c_k$ . The remaining divisions compute

$$p'_{n_1, \dots, n_{k-1}}(p_{n_1, \dots, n_{k-1}}(c_1, \dots, c_{k-1})),$$

which equals  $(c_1, \dots, c_{k-1})$  by induction hypothesis.  $\square$

**5.12. Example.** Suppose  $(n_1, n_2, n_3, n_4, n_5) = (4, 6, 5, 4, 2)$ , so  $n = n_1 n_2 n_3 n_4 n_5 = 960$ . Then

$$p(3, 1, 0, 2, 1) = 3 \cdot (6 \cdot 5 \cdot 4 \cdot 2) + 1 \cdot (5 \cdot 4 \cdot 2) + 0 \cdot (4 \cdot 2) + 2 \cdot (2) + 1 = 765.$$

To compute  $p^{-1}(222)$ , first divide 222 by  $n_5 = 2$  to get  $q_5 = 111$  and  $r_5 = 0$ . Then divide 111 by  $n_4 = 4$  to get  $q_4 = 27$  and  $r_4 = 3$ . Then divide 27 by  $n_3 = 5$  to get  $q_3 = 5$  and  $r_3 = 2$ . Then divide 5 by  $n_2 = 6$  to get  $q_2 = 0$  and  $r_2 = 5$ . Finally, divide 0 by  $n_1 = 4$  to get  $q_1 = 0$  and  $r_1 = 0$ . We conclude that  $p^{-1}(222) = (0, 5, 2, 3, 0)$ .

**5.13. Remark.** If  $n_1 = \dots = n_k = b$ , then  $p_{b, b, \dots, b}^{-1}(z)$  computes the base- $b$  expansion of  $z$  (for  $0 \leq z < b^k$ ). Here,  $p$  and  $p^{-1}$  are bijections between the set of words  $\{0, 1, \dots, b-1\}^k$  and the set of integers  $\{0, 1, 2, \dots, b^k - 1\}$ . Taking  $b = 10$ , we see that the decimal representation of positive integers is a special case of the bijective product rule.

**5.14. Remark.** Here is another algorithm for computing  $p_{n_1, \dots, n_k}^{-1}(m) = p^{-1}(p(c_1, \dots, c_k))$  that recovers the numbers  $(c_1, \dots, c_k)$  from left to right. First, divide  $m$  by  $n_2 n_3 \dots n_k$  to obtain a quotient  $q_1$  and a remainder  $r_1$ . Set  $c_1 = q_1$ , and recover  $(c_2, \dots, c_k)$  by recursively computing  $p_{n_2, \dots, n_k}^{-1}(r_1)$  in the same fashion. We let the reader prove that this algorithm does implement the inverse of  $p_{n_1, \dots, n_k}$ .

We can now describe the general strategy for converting informal counting arguments based on the product rule to ranking and unranking algorithms. When using the product rule, we uniquely construct objects in a set  $S$  by making  $k$  choices, such that there are always  $n_i$  ways to make the  $i$ th choice. There is often a natural ordering of the choices that can be made at the  $i$ th stage, given the choices that have already been made; thus we can number the available choices at this stage  $0, 1, \dots, n_i - 1$ . Then the informal construction process for manufacturing objects in  $S$  can be translated into a formal bijection  $f : \underline{n}_1 \times \dots \times \underline{n}_k \rightarrow S$ , where  $f(c_1, \dots, c_k)$  is the object constructed by making the choice numbered  $c_i$  at the  $i$ th stage, for  $1 \leq i \leq k$ . To solve the unranking problem for  $S$ , we use the composite bijection

$$f \circ p_{n_1, \dots, n_k}^{-1} : \underline{n} \rightarrow S.$$

Similarly, if we can find an algorithm that recovers the choices  $c_i$  from the final object  $x \in S$ , then we can compute  $f^{-1}$ . The ranking problem for  $S$  is then solved by the bijection

$$p_{n_1, \dots, n_k} \circ f^{-1} : S \rightarrow \underline{\mathbf{n}}.$$

In the next several sections, we apply this idea to find ranking and unranking bijections for many commonly occurring families of combinatorial objects.

## 5.4 Ranking Words

**5.15. Example: Four-Letter Words.** Let  $S$  be the set of all four-letter words. We can build elements of  $S$  by choosing the first letter, then the second, third, and fourth letters. To describe this choice process formally as a bijection, let  $A = \{a, b, \dots, z\}$  be the alphabet. We identify a word  $w = w_1 w_2 w_3 w_4 \in S$  with the 4-tuple  $(w_1, w_2, w_3, w_4) \in A \times A \times A \times A$ . Choose a fixed bijection  $f : A \rightarrow \underline{\mathbf{26}}$ ; for instance, we can use the standard alphabetical ordering given by

$$f(a) = 0, f(b) = 1, f(c) = 2, \dots, f(y) = 24, f(z) = 25.$$

Then  $f \times f \times f \times f$  is a bijection from  $S = A \times A \times A \times A$  to  $\underline{\mathbf{26}}^4$ . Composing with the bijection  $p : \underline{\mathbf{26}} \times \underline{\mathbf{26}} \times \underline{\mathbf{26}} \times \underline{\mathbf{26}} \rightarrow \underline{\mathbf{26}}^4 = \underline{\mathbf{456,976}}$ , we obtain a *ranking map*  $r : S \rightarrow \underline{\mathbf{456,976}}$ . For example,

$$r(\text{goop}) = p_{26,26,26,26}(6, 14, 14, 15) = 6 \cdot 26^3 + 14 \cdot 26^2 + 14 \cdot 26^1 + 15 = 115,299;$$

$$r(\text{pogo}) = p_{26,26,26,26}(15, 14, 6, 14) = 15 \cdot 26^3 + 14 \cdot 26^2 + 6 \cdot 26^1 + 14 = 273,274.$$

The inverse of  $r$  is the *unranking map*  $u : \underline{\mathbf{456,976}} \rightarrow S$ . To compute  $u(x)$ , we first express  $x$  in base 26 (this is what  $p_{26,26,26,26}^{-1}$  does), and then use the inverse of  $f$  to convert back to letters. For example,

$$u(200,000) = u(11 \cdot 26^3 + 9 \cdot 26^2 + 22 \cdot 26 + 8) = f^{-1}(11)f^{-1}(9)f^{-1}(22)f^{-1}(8) = \text{ljwi}.$$

In general, if  $A$  is an  $m$ -letter alphabet, we can rank the set of  $k$ -letter words  $A^k$  by fixing a bijection  $f : A \rightarrow \underline{\mathbf{m}}$  and computing

$$r(w_1 w_2 \cdots w_k) = p_{m,m,\dots,m}(f(w_1), \dots, f(w_k)).$$

To unrank an integer  $z \in \underline{\mathbf{m}}^k$ , write  $z = d_{k-1} \cdots d_0$  in base  $m$  and then replace each digit  $d_i$  by  $f^{-1}(d_i)$ .

**5.16. Example: Three-Letter Words.** Now consider  $S = \{a, b, c\}^3$ . Define  $f : \{a, b, c\} \rightarrow \underline{\mathbf{3}}$  by  $f(a) = 0$ ,  $f(b) = 1$ , and  $f(c) = 2$ . The ranking map  $r : S \rightarrow \underline{\mathbf{27}}$  is defined by

$$r(w_1 w_2 w_3) = f(w_1) \cdot 9 + f(w_2) \cdot 3 + f(w_3).$$

To define the unranking map  $u : \underline{\mathbf{27}} \rightarrow S$ , write  $z \in \underline{\mathbf{27}}$  as  $z = d_2 d_1 d_0$  in base 3. Then

$$u(z) = u(d_2 d_1 d_0) = f^{-1}(d_2)f^{-1}(d_1)f^{-1}(d_0).$$

These bijections set up the following one-to-one correspondences between  $S$ ,  $\underline{\mathbf{3}} \times \underline{\mathbf{3}} \times \underline{\mathbf{3}}$ , and  $\underline{\mathbf{27}}$ :

aaa $\leftrightarrow$ 000=0	baa $\leftrightarrow$ 100=9	caa $\leftrightarrow$ 200=18
aab $\leftrightarrow$ 001=1	bab $\leftrightarrow$ 101=10	cab $\leftrightarrow$ 201=19
aac $\leftrightarrow$ 002=2	bac $\leftrightarrow$ 102=11	cac $\leftrightarrow$ 202=20
aba $\leftrightarrow$ 010=3	bba $\leftrightarrow$ 110=12	cba $\leftrightarrow$ 210=21
abb $\leftrightarrow$ 011=4	bbb $\leftrightarrow$ 111=13	cbb $\leftrightarrow$ 211=22
abc $\leftrightarrow$ 012=5	bbc $\leftrightarrow$ 112=14	cbc $\leftrightarrow$ 212=23
aca $\leftrightarrow$ 020=6	bca $\leftrightarrow$ 120=15	cca $\leftrightarrow$ 220=24
acb $\leftrightarrow$ 021=7	ccb $\leftrightarrow$ 121=16	ccb $\leftrightarrow$ 221=25
acc $\leftrightarrow$ 022=8	bcc $\leftrightarrow$ 122=17	ccc $\leftrightarrow$ 222=26

Notice that, as  $z$  runs from 0 to 26, the words in  $S$  are generated in alphabetical order.

More generally, using the bijective product rule will generate the elements of a set  $S$  in a certain *lexicographic order* that is determined by the nature of the “choice bijection”  $\mathbf{n}_1 \times \mathbf{n}_2 \times \cdots \times \mathbf{n}_k \rightarrow S$  used to build the objects in  $S$  from a sequence of choices. According to our definition of  $p_{n_1, \dots, n_k}$  and its inverse, the first choice in the sequence (which can occur in  $n_1$  ways) is deemed “most significant,” and the last choice (which can occur in  $n_k$  ways) is “least significant.” If we unrank  $0, 1, \dots, n-1$  in this order, we obtain a list that begins with the  $n_2 \cdots n_k$  objects that can be made by choosing zero in the first choice. Next we get all the objects that can be made by choosing one in the first choice, etc. Each such sublist is also arranged lexicographically, according to the choices made at stages  $2, 3, \dots, k$ .

**5.17. Example: Words with Restrictions.** Let  $S$  be the set of four-letter words  $w_1 w_2 w_3 w_4$  that begin and end with consonants and have a vowel in the second position. Choosing letters from left to right and using the product rule, we see that  $|S| = 21 \cdot 5 \cdot 26 \cdot 21 = 57,330$ . Let  $C$ ,  $V$ , and  $A$  denote the set of consonants, vowels, and all letters, respectively. The usual alphabetical order defines bijections  $C \rightarrow \underline{21}$ ,  $V \rightarrow \underline{5}$ , and  $A \rightarrow \underline{26}$ ; for example,

$$V \rightarrow \underline{5} \text{ via } a \mapsto 0, e \mapsto 1, i \mapsto 2, o \mapsto 3, u \mapsto 4.$$

We obtain a ranking map  $r : S \rightarrow \underline{57,330}$  by defining

$$r(w_1 w_2 w_3 w_4) = p_{21,5,26,21}(w'_1, w'_2, w'_3, w'_4),$$

where  $w'_i$  denotes the image of  $w_i$  under the appropriate bijection. For example,

$$r(\text{host}) = p_{21,5,26,21}(5, 3, 18, 15) = 5 \cdot (5 \cdot 26 \cdot 21) + 3 \cdot (26 \cdot 21) + 18 \cdot (21) + 15 = 15,681.$$

We unrank by applying  $p_{21,5,26,21}^{-1}$  and then decoding to letters. For example, repeated division shows that

$$p_{21,5,26,21}^{-1}(44001) = (16, 0, 15, 6),$$

and therefore  $u(44001) = \text{vapj}$ . This unranking method generates the words in  $S$  in alphabetical order.

**5.18. Example: License Plates.** A California license plate consists of a digit, followed by three letters, followed by three digits. We can use the preceding ideas to rank and unrank license plates. For instance,

$$r(3\text{PZY}292) = p_{10,26,26,26,1000}(3, 15, 25, 24, 292) = 63,542,292.$$

## 5.5 Ranking Permutations

In the examples considered so far, the choices made at each stage of the product rule did not depend on what choices were made in previous stages. This section studies the more complicated situation where the available choices do depend on what happened earlier. We illustrate this situation by solving the ranking and unranking problems for permutations.

Suppose  $A$  is an  $n$ -letter alphabet. Recall that a  $k$ -permutation of  $A$  is a word  $w = w_1 w_2 \cdots w_k$ , where the  $w_i$ 's are *distinct* elements of  $A$ . Let  $S$  be the set of all  $k$ -permutations of  $A$ . Using the ordinary product rule, we build elements  $w$  in  $S$  by choosing  $w_1 \in A$  in  $n$  ways, then choosing  $w_2 \in A \sim \{w_1\}$  in  $n - 1$  ways, and so on. At the  $i$ th stage (where  $1 \leq i \leq k$ ), we choose  $w_i \in A \sim \{w_1, w_2, \dots, w_{i-1}\}$  in  $n - (i - 1)$  ways. Thus,  $|S| = n(n - 1) \cdots (n - k + 1) = (n)_{\downarrow k}$ . Notice that the set of choices available at the  $i$ th stage depends on the choices made earlier, but the *cardinality* of this set is independent of previous choices. (This last fact is a key hypothesis of the product rule.)

Let us rephrase the preceding counting argument to obtain a bijection between  $S$  and the product set  $\underline{n} \times \underline{n-1} \times \cdots \times \underline{n-k+1}$ . Fix a total ordering  $x = (x_0, x_1, \dots, x_{n-1})$  of the letters in  $A$ ; equivalently, fix a bijection  $x : \underline{n} \rightarrow A$ . Suppose  $w = w_1 w_2 \cdots w_k \in S$ . We must map  $w$  to a  $k$ -tuple  $(j_1, j_2, \dots, j_k)$ , where  $0 \leq j_i < n - (i - 1)$ . To compute  $j_1$ , locate  $w_1$  in the sequence  $(x_0, x_1, \dots, x_{n-1})$ , let  $j_1$  be the number of letters preceding  $w_1$  in the sequence, and then erase  $w_1$  from the sequence to get a new sequence  $x'$ . To compute  $j_2$ , find  $w_2$  in the sequence  $x'$ , let  $j_2$  be the number of letters preceding it, and then erase  $w_2$  to get a new sequence  $x''$ . Continue similarly to generate the remaining  $j_i$ 's. This process is reversible, as demonstrated in the next example, so we have defined the desired bijection. Combining this bijection (and its inverse) with the maps  $p_{n,n-1,\dots,n-k+1}$  and  $p_{n,n-1,\dots,n-k+1}^{-1}$ , we obtain the desired ranking and unranking maps. One may verify that these maps correspond to the alphabetic ordering of permutations specified by the given total ordering of the alphabet  $A$ .

**5.19. Example.** Let  $n = 8$ ,  $k = 5$ , and  $A = (a,b,c,d,e,f,g,h)$  with the usual alphabetical ordering. Let  $w = \text{cfbgd} \in S$ . We compute  $(j_1, \dots, j_5)$  as follows:

$$\begin{array}{ll} 2 \text{ letters precede } c \text{ in } (a,b,c,d,e,f,g,h), \text{ so} & j_1 = 2; \\ 4 \text{ letters precede } f \text{ in } (a,b,d,e,f,g,h), \text{ so} & j_2 = 4; \\ 1 \text{ letter precedes } b \text{ in } (a,b,d,e,g,h), \text{ so} & j_3 = 1; \\ 3 \text{ letters precede } g \text{ in } (a,d,e,g,h), \text{ so} & j_4 = 3; \\ 1 \text{ letter precedes } d \text{ in } (a,d,e,h), \text{ so} & j_5 = 1. \end{array}$$

Thus,  $\text{cfbgd} \mapsto (2, 4, 1, 3, 1)$ . The rank of this word is therefore

$$p_{8,7,6,5,4}(2, 4, 1, 3, 1) = 2 \cdot (7 \cdot 6 \cdot 5 \cdot 4) + 4 \cdot (6 \cdot 5 \cdot 4) + 1 \cdot (5 \cdot 4) + 3 \cdot (4) + 1 = 2193.$$

Next, let us unrank the integer 982. First, repeated division gives

$$p_{8,7,6,5,4}^{-1}(982) = (1, 1, 1, 0, 2).$$

Since  $j_1 = 1$ , the first letter of the desired word must be  $b$ . Removing  $b$  from the alphabet gives  $(a,c,d,e,f,g,h)$ . Since  $j_2 = 1$ , the second letter of the desired word is  $c$ . Removing  $c$  from the previous list gives  $(a,d,e,f,g,h)$ . Continuing in this way, we see that 982 unranks to give the word  $\text{bcdag}$ .

**5.20. Example.** Let  $S$  be the set of permutations of  $(1, 2, 3, 4, 5, 6)$ . Using the procedure above to rank the permutation  $(4, 6, 2, 1, 5, 3)$ , we first compute  $(j_1, \dots, j_6) = (3, 4, 1, 0, 1, 0)$  and then calculate  $p_{6,5,4,3,2,1}(3, 4, 1, 0, 1, 0) = 463$ . To unrank the integer 397, first calculate  $p_{6,5,4,3,2,1}^{-1}(397) = (3, 1, 2, 0, 1, 0)$ . Then use these position numbers to recover the permutation  $(4, 2, 5, 1, 6, 3)$ .

## 5.6 Ranking Subsets

In 1.42, we used the product rule to prove that the number of  $k$ -element subsets of an  $n$ -element set is  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ . This enumeration result was obtained *indirectly*, by enumerating  $k$ -permutations of an  $n$ -element set in two ways and then dividing the resulting equation by  $k!$ . In general, the operation of division presents serious problems when attempting to construct bijections. Therefore, we will adopt a different approach to the problem of ranking and unranking subsets. Instead of using the bijective product rule, we will apply the bijective sum rule to the recursion characterizing the binomial coefficients. This will lead us to recursive algorithms for ranking and unranking subsets.

For convenience, write  $C(n, k)$  for the number of  $k$ -element subsets of an  $n$ -element set. In 2.25, we saw that these numbers satisfy the recursion

$$C(n, k) = C(n-1, k) + C(n-1, k-1) \quad (0 < k \leq n) \quad (5.1)$$

with initial conditions  $C(n, 0) = 1$ . This recursion came from a combinatorial argument involving the sum rule. Using the bijective sum rule instead will lead directly to recursively defined bijections for ranking and unranking. For each alphabet  $A$ , introduce the temporary notation  $S_k(A)$  to denote the set of all  $k$ -element subsets of  $A$ . We assume that all alphabets to be considered are equipped with some fixed total ordering that allows us to rank and unrank individual letters of the alphabet. Suppose  $A = (x_0, x_1, \dots, x_{n-1})$  is such an alphabet with  $n$  letters. We can write  $S_k(A)$  as the disjoint union of sets  $T$  and  $U$ , where  $T$  consists of all subsets that do not contain  $x_{n-1}$  and  $U$  consists of all subsets that contain  $x_{n-1}$ . Note that  $T = S_k(A \sim \{x_{n-1}\})$ , and  $U$  corresponds to  $S_{k-1}(A \sim \{x_{n-1}\})$  via a bijection that deletes  $x_{n-1}$  from a subset belonging to  $U$ . We can use recursion to obtain ranking and unranking maps for  $S_k(A \sim \{x_{n-1}\})$  and  $S_{k-1}(A \sim \{x_{n-1}\})$ , as these involve subsets drawn from smaller alphabets. Then we combine these maps using the bijective sum rule to get ranking and unranking maps for  $S_k(A)$ .

Writing out the definitions, we arrive at the following recursive ranking algorithm for mapping a subset  $B \in S_k(A)$  to an integer:

- If  $k = 0$  (so  $B = \emptyset$ ), then return the answer 0.
- If  $k > 0$  and the last letter  $x$  in  $A$  does not belong to  $B$ , then return the ranking of  $B$  relative to the set  $S_k(A \sim \{x\})$ , which we compute recursively using this very algorithm.
- If  $k > 0$  and the last letter  $x$  in  $A$  does belong to  $B$ , let  $i$  be the ranking of  $B' = B \sim \{x\}$  relative to the set  $S_{k-1}(A \sim \{x\})$  (computed recursively), and return the answer  $i + C(n-1, k)$ . Note that  $C(n-1, k)$  can be computed using the recursion (5.1) for binomial coefficients.

The inverse map is the following recursive unranking algorithm that maps an integer  $m$  to a subset  $B \in S_k(A)$ :

- If  $k = 0$  (so  $m$  must be zero), then return  $\emptyset$ .
- If  $k > 0$  and  $0 \leq m < C(n-1, k)$ , then return the result of unranking  $m$  relative to the set  $S_k(A \sim \{x\})$ , where  $x$  is the last letter of  $A$ .
- If  $k > 0$  and  $C(n-1, k) \leq m < C(n, k)$ , then let  $B'$  be the unranking of  $m - C(n-1, k)$  relative to the set  $S_{k-1}(A \sim \{x\})$ , and return  $B' \cup \{x\}$ .

**5.21. Example.** Let  $A = (a, b, c, d, e, f, g, h)$ , and let us rank the subset  $B = \{c, d, f, g\} \in S_4(A)$ . Since  $h \notin B$ , we recursively proceed to rank  $B$  relative to the 7-letter alphabet  $A_1 = (a, b, c, d, e, f, g)$ . The new last letter  $g$  belongs to  $B$ , so we must add  $C(7 - 1, 4) = 15$  to the rank of  $B_1 = \{c, d, f\}$  relative to  $(a, b, c, d, e, f)$ . The last letter  $f$  belongs to  $B_1$ , so we must add  $C(6 - 1, 3) = 10$  to the rank of  $B_2 = \{c, d\}$  relative to  $(a, b, c, d, e)$ . This is the same as the rank of  $B_2$  relative to  $(a, b, c, d)$ , which is  $C(4 - 1, 2) = 3$  plus the rank of  $\{c\}$  relative to  $(a, b, c)$ , which is  $C(3 - 1, 1) = 2$  plus the rank of  $\emptyset$  relative to  $(a, b)$ . Two more reductions reveal that the latter rank is zero. Adding up the contributions, we see that the rank of  $B$  is

$$\binom{3-1}{1} + \binom{4-1}{2} + \binom{6-1}{3} + \binom{7-1}{4} = 30.$$

Generalizing the pattern in the previous example, we can give the following non-recursive formula for the rank of a subset.

**5.22. Sum Formula for Ranking Subsets.** If  $A = (x_0, x_1, \dots, x_{n-1})$  and  $B = \{x_{i_1}, x_{i_2}, \dots, x_{i_k}\}$  where  $i_1 < i_2 < \dots < i_k$ , then the rank of  $B$  relative to  $S_k(A)$  is  $\sum_{j=1}^k \binom{i_j}{j}$ .

A routine induction argument can be used to prove this formula formally.

**5.23. Example.** Now we illustrate the recursive unranking algorithm. Let us unrank the integer 53 to obtain an object  $B \in S_4(A)$ , where  $A = (a, b, c, d, e, f, g, h)$ . Here  $n = 8$  and  $k = 4$ . Since  $C(7, 4) = 35 \leq 53$ , we know that  $h \in B$ . We proceed by unranking  $53 - 35 = 18$  to get a 3-element subset of  $(a, b, c, d, e, f, g)$ . Now  $C(6, 3) = 20 > 18$ , so  $g$  does not lie in the subset. We proceed to unrank 18 to get a 3-element subset of  $(a, b, c, d, e, f)$ . Now  $C(5, 3) = 10 \leq 18$ , so  $f$  does belong to  $B$ . We continue, unranking  $18 - 10 = 8$  to get a two-element subset of  $(a, b, c, d, e)$ . Since  $C(4, 2) = 6 \leq 8$ ,  $e \in B$  and we continue by unranking 2. We have  $C(3, 1) = 3 > 2$ , so  $d \notin B$ . But at the next stage  $C(2, 1) = 2 \leq 2$ , so  $c \in B$ . We conclude, finally, that  $B = \{c, e, f, h\}$ .

As before, we can describe this algorithm iteratively instead of recursively.

**5.24. Unranking Algorithm for Subsets.** Suppose  $A = (x_0, x_1, \dots, x_{n-1})$  and we are unranking an integer  $m$  to get a  $k$ -element subset  $B$  of  $A$ . Repeatedly perform the following steps until  $k$  becomes zero: let  $i$  be the largest integer such that  $C(i, k) \leq m$ ; declare that  $x_i \in B$ ; replace  $m$  by  $m - C(i, k)$  and decrement  $k$  by 1.

We close with a remark about the ordering of subsets associated to the ranking and unranking algorithms described above. Let  $x$  be the last letter of  $A$ . If we unrank the integers  $0, 1, 2, \dots$  in this order to obtain a listing of  $S_k(A)$ , we will obtain all  $k$ -element subsets of  $A$  not containing  $x$  first, and all  $k$ -element subsets of  $A$  containing  $x$  second. Each of these sublists is internally ordered in the same way according to the next-to-last letter of  $A$ , and so on recursively. In contrast, if we had used the bijective sum rule on the recursion

$$C(n, k) = C(n - 1, k - 1) + C(n - 1, k)$$

(in which the order of the summands is swapped), then the ordering rules at each level of this hierarchy would be reversed. Similarly, the reader can construct variant ranking algorithms in which the first letter of the alphabet is considered “most significant,” etc. Some of these variants are explored in the exercises.

## 5.7 Ranking Anagrams

Next we study the problem of ranking and unranking anagrams. Recall that  $\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  is the set of all words of length  $n = n_1 + \cdots + n_k$  consisting of  $n_i$  copies of  $a_i$  for  $1 \leq i \leq k$ . We have seen (§1.9) that these sets are counted by the multinomial coefficients:

$$|\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})| = \binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}.$$

There are at least three ways of deriving this formula. One way counts permutations of  $n$  distinct letters in two ways, and solves for the number of anagrams by division. This method is not easily converted into a ranking algorithm. A second way uses the product rule, choosing the positions for the  $n_1$  copies of  $a_1$ , then the positions for the  $n_2$  copies of  $a_2$ , and so on. Combining the bijective product rule with the ranking algorithm for subsets presented earlier, this method does lead to a ranking algorithm for anagrams. A third way to count anagrams involves finding recursions involving the multinomial coefficients (§2.5). This is the approach we will pursue here.

Let  $C(n; n_1, \dots, n_k)$  be the number of rearrangements of  $n$  letters, where there are  $n_i$  letters of type  $i$ . Classifying words by their first letter leads to the recursion

$$C(n; n_1, \dots, n_k) = \sum_{i=1}^k C(n-1; n_1, \dots, n_i-1, \dots, n_k).$$

Applying the bijective sum rule to this recursion, we are led to recursive ranking and unranking algorithms for anagrams.

Here are the details of the algorithms. We recursively define ranking maps

$$r = r_{a_1^{n_1} \cdots a_k^{n_k}} : \mathcal{R}(a_1^{n_1} \cdots a_k^{n_k}) \rightarrow \underline{\mathbf{m}},$$

where  $m = (n_1 + \cdots + n_k)! / (n_1! \cdots n_k!)$ . If any  $n_i$  is negative,  $r$  is the function with graph  $\emptyset$ . If all  $n_i$ 's are zero,  $r$  is the function sending the empty word to 0. To compute  $r(w)$  in the remaining case, suppose  $a_i$  is the first letter of  $w$ . Write  $w = a_i w'$ . Return the answer

$$r(w) = \sum_{j < i} C(n-1; n_1, \dots, n_j-1, \dots, n_k) + r_{a_1^{n_1} \cdots a_i^{n_i-1} \cdots a_k^{n_k}}(w'),$$

where  $r(w')$  is computed recursively by the same algorithm.

Next, we define the corresponding unranking maps

$$u = u_{a_1^{n_1} \cdots a_k^{n_k}} : \underline{\mathbf{m}} \rightarrow \mathcal{R}(a_1^{n_1} \cdots a_k^{n_k}).$$

Use the only possible maps if some  $n_i < 0$  or if all  $n_i = 0$ . Otherwise, to unrank  $s \in \underline{\mathbf{m}}$ , first find the maximal index  $i$  such that  $n_i > 0$  and  $\sum_{j < i} C(n-1; n_1, \dots, n_j-1, \dots, n_k) \leq s$ ; let  $s'$  be the difference between  $s$  and this sum. Recursively compute

$$w' = u_{a_1^{n_1} \cdots a_i^{n_i-1} \cdots a_k^{n_k}}(s');$$

finally, return the answer  $w = a_i w'$ . This unranking algorithm induces a listing of the anagrams in  $\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  “in alphabetical order” relative to the alphabet ordering  $a_1 < a_2 < \cdots < a_k$ .

**5.25. Example.** Let us compute the rank of the word  $w = \text{abbcacb}$  in  $\mathcal{R}(a^2b^3c^2)$ ; here  $n = 7$ ,  $n_1 = 2$ ,  $n_2 = 3$ , and  $n_3 = 2$ . Erasing the first letter  $a$ , we see that the rank of  $w$  equals zero plus the rank of  $w_1 = \text{bbcacb}$ ; now  $n = 6$ ,  $n_1 = 1$ ,  $n_2 = 3$ , and  $n_3 = 2$ . Erasing  $b$ , we must now add  $\binom{5}{0,3,2} = 10$  to the rank of  $w_2 = \text{bcacb}$ ; now  $n = 5$ ,  $n_1 = 1$ ,  $n_2 = 2$ , and  $n_3 = 2$ . Erasing the next  $b$ , we must add  $\binom{4}{0,2,2} = 6$  to the rank of  $w_3 = \text{cacb}$ ; now  $n = 4$ ,  $n_1 = 1$ ,  $n_2 = 1$ , and  $n_3 = 2$ . Erasing  $c$ , we must add  $\binom{3}{0,1,2} + \binom{3}{1,0,2} = 6$  to the rank of  $w_4 = \text{acb}$ ; now  $n = 3$ ,  $n_1 = 1$ ,  $n_2 = 1$ , and  $n_3 = 1$ . Continuing in this way, one sees that the rank of  $\text{acb}$  is 1. Thus, the rank of the original word is  $10 + 6 + 6 + 1 = 23$ .

Next, let us unrank 91 to obtain a word  $w$  in  $\mathcal{R}(a^2b^3c^2)$ . To determine the first letter of  $w$ , note that  $0 \leq 91$ ,  $\binom{6}{1,3,2} = 60 \leq 91$ , but  $\binom{6}{1,3,2} + \binom{6}{2,2,2} = 150 > 91$ . Thus, the first letter is  $b$ , and we continue by unranking  $91 - 60 = 31$  to obtain a word in  $\mathcal{R}(a^2b^2c^2)$ . This time, we have  $0 \leq 31$ ,  $\binom{5}{1,2,2} = 30 \leq 31$ , but  $\binom{5}{1,2,2} + \binom{5}{2,1,2} = 60 > 31$ . So the second letter is  $b$ , and we continue by unranking  $31 - 30 = 1$  to obtain a word in  $\mathcal{R}(a^2b^1c^2)$ . It is routine to check that the next two letters are both  $a$ , and we continue by unranking 1 to obtain a word in  $\mathcal{R}(b^1c^2)$ . The word we get is  $\text{cbc}$ , so the unranking of 91 is the word  $\text{bbaacbc}$ .

## 5.8 Ranking Integer Partitions

In this section, we devise ranking and unranking algorithms for integer partitions by applying the bijective sum rule to the recursion 2.42. Let  $P(n, k)$  be the set of integer partitions of  $n$  with largest part  $k$ , and let  $p(n, k) = |P(n, k)|$ . Recall from 2.42 that these numbers satisfy

$$p(n, k) = p(n - k, k) + p(n - 1, k - 1) \quad (n, k > 0).$$

The first term on the right counts elements of  $P(n, k)$  in which the largest part occurs at least twice (deleting the first copy of this part gives a bijection onto  $P(n - k, k)$ ). The second term on the right counts elements of  $P(n, k)$  in which the largest part occurs exactly once (reducing this part by one gives a bijection onto  $P(n - 1, k - 1)$ ). Combining these bijections with the bijective sum rule, we obtain recursively determined ranking maps  $r = r_{n,k} : P(n, k) \rightarrow \mathbf{p}(\mathbf{n}, \mathbf{k})$ . To find  $r_{n,k}(\mu)$ , consider three cases. If  $\mu$  has only one part (which happens when  $n = k$ ), return 0. If  $k = \mu_1 = \mu_2$ , return  $r_{n-k,k}((\mu_2, \mu_3, \dots))$ . If  $k = \mu_1 > \mu_2$ , return  $p(n - k, k) + r_{n-1,k-1}((\mu_1 - 1, \mu_2, \dots))$ . The unranking maps  $u = u_{n,k} : \mathbf{p}(\mathbf{n}, \mathbf{k}) \rightarrow P(n, k)$  operate as follows. To compute  $u(m)$  where  $0 \leq m < p(n, k)$ , consider two cases. If  $0 \leq m < p(n - k, k)$ , recursively compute  $\nu = u_{n-k,k}(m)$  and return the answer  $\mu = (k, \nu_1, \nu_2, \dots)$ . If  $p(n - k, k) \leq m < p(n, k)$ , recursively compute  $\nu = u_{n-1,k-1}(m - p(n - k, k))$  and return the answer  $\mu = (\nu_1 + 1, \nu_2, \nu_3, \dots)$ .

**5.26. Example.** Let us compute  $r_{8,3}(\mu)$ , where  $\mu = (3, 3, 1, 1)$ . Since  $\mu_1 = \mu_2$ , the rank will be  $r_{5,3}(\nu)$ , where  $\nu = (3, 1, 1)$ . Next, since  $\nu_1 \neq \nu_2$ , we have

$$r_{5,3}(3, 1, 1) = p(2, 3) + r_{4,2}(2, 1, 1) = r_{4,2}(2, 1, 1).$$

The first two parts of the new partition are again different, so

$$r_{4,2}(2, 1, 1) = p(2, 2) + r_{3,1}(1, 1, 1) = 1 + r_{3,1}(1, 1, 1).$$

After several more steps, we find that  $r_{3,1}(1, 1, 1) = 0$ , so  $r_{8,3}(\mu) = 1$ . Thus  $\mu$  is the second partition in the listing of  $P(8, 3)$  implied by the ranking algorithm; the first partition in this list, which has rank 0, is  $(3, 3, 2)$ .



Next, let us compute  $\mu = u_{10,4}(6)$ . First,  $p(6, 4) = 2 \leq 6$ , so  $\mu$  will be obtained by adding one to the first part of  $\nu = u_{9,3}(4)$ . Second,  $p(6, 3) = 3 \leq 4$ , so  $\nu$  will be obtained by adding one to the first part of  $\rho = u_{8,2}(1)$ . Third,  $p(6, 2) = 3 > 1$ , so  $\rho$  will be obtained by adding a new first part of length 2 to  $\xi = u_{6,2}(1)$ . Fourth,  $p(4, 2) = 2 > 1$ , so  $\xi$  will be obtained by adding a new first part of length 2 to  $\zeta = u_{4,2}(1)$ . Fifth,  $p(2, 2) = 1 \leq 1$ , so  $\zeta$  will be obtained by adding one to the first part of  $\omega = u_{3,1}(0)$ . We must have  $\omega = (1, 1, 1)$ , this being the unique element of  $P(3, 1)$ . Working our way back up the chain, we successively find that

$$\zeta = (2, 1, 1), \quad \xi = (2, 2, 1, 1), \quad \rho = (2, 2, 2, 1, 1), \quad \nu = (3, 2, 2, 1, 1),$$

and finally  $\mu = u_{10,4}(6) = (4, 2, 2, 1, 1)$ .

Now that we have algorithms to rank and unrank the sets  $P(n, k)$ , we can apply the bijective sum rule to the identity

$$p(n) = p(n, n) + p(n, n-1) + \cdots + p(n, 1)$$

to rank and unrank the set  $P(n)$  of all integer partitions of  $n$ .

**5.27. Example.** Let us enumerate all the integer partitions of 6. We obtain this list of partitions by concatenating the lists associated to the sets

$$P(6, 6), P(6, 5), \dots, P(6, 1),$$

written in this order. In turn, each of these lists can be constructed by applying the unranking maps  $u_{6,k}$  to the integers  $0, 1, 2, \dots, p(6, k) - 1$ . The reader can verify that this procedure leads to the following list:

$$\begin{aligned} (6), (5, 1), (4, 2), (4, 1, 1), (3, 3), (3, 2, 1), (3, 1, 1, 1), \\ (2, 2, 2), (2, 2, 1, 1), (2, 1, 1, 1, 1), (1, 1, 1, 1, 1, 1). \end{aligned}$$

One may also check that the list obtained in this way presents the integer partitions of  $n$  in decreasing lexicographic order (as defined in 10.36).

## 5.9 Ranking Set Partitions

Next, we consider the ranking and unranking of set partitions (which are counted by Stirling numbers of the second kind and Bell numbers). The recursion for Stirling numbers involves both addition and multiplication, so our recursive algorithms will use both the bijective sum rule and the bijective product rule.

Let  $SP(n, k)$  be the set of all set partitions of  $\{1, 2, \dots, n\}$  into exactly  $k$  blocks, and let  $S(n, k) = |SP(n, k)|$  be the associated Stirling number of the second kind. Recall from 2.52 that

$$S(n, k) = S(n-1, k-1) + kS(n-1, k) \quad (n, k > 0).$$

The first term counts set partitions in  $SP(n, k)$  such that  $n$  is in a block by itself; removal of this block gives a bijection onto  $SP(n-1, k-1)$ . The second term counts set partitions  $\pi$  in  $SP(n, k)$  such that  $n$  belongs to a block with other elements. Starting with any set partition  $\pi'$  in  $SP(n-1, k)$ , we can build such a set partition  $\pi \in SP(n, k)$  by adding  $n$

to any of the  $k$  nonempty blocks of  $\pi'$ . We number the blocks of  $\pi'$  using  $0, 1, \dots, k-1$  by arranging the minimum elements of these blocks in increasing order. For example, if  $\pi' = \{\{6, 3, 5\}, \{2\}, \{1, 7\}, \{8, 4\}\}$ , then block 0 of  $\pi'$  is  $\{1, 7\}$ , block 1 is  $\{2\}$ , block 2 is  $\{3, 5, 6\}$ , and block 3 is  $\{4, 8\}$ .

The ranking maps  $r = r_{n,k} : SP(n, k) \rightarrow \mathbf{S}(\mathbf{n}, \mathbf{k})$  are defined recursively as follows. Use the only possible maps if  $k \leq 0$  or  $k > n$ . For  $0 < k \leq n$ , compute  $r_{n,k}(\pi)$  as follows. If  $\{n\} \in \pi$ , return the answer  $r_{n-1,k-1}(\pi \sim \{\{n\}\})$ . Otherwise, let  $\pi'$  be obtained from  $\pi$  by deleting  $n$  from whatever block contains it, and let  $i$  be the index of the block of  $\pi'$  that used to contain  $n$ . Return  $S(n-1, k-1) + p_{k,S(n-1,k)}(i, r_{n-1,k}(\pi'))$ .

Similarly, we define the unranking maps  $u = u_{n,k} : \mathbf{S}(\mathbf{n}, \mathbf{k}) \rightarrow SP(n, k)$  as follows. Assume  $n, k > 0$  and we are computing  $u_{n,k}(m)$ . If  $0 \leq m < S(n-1, k-1)$ , then return  $u_{n-1,k-1}(m) \cup \{\{n\}\}$ . If  $S(n-1, k-1) \leq m < S(n, k)$ , first compute  $(i, j) = p_{k,S(n-1,k)}^{-1}(m - S(n-1, k-1))$ . Next, calculate the partition  $\pi' = u_{n-1,k}(j)$  by unranking  $j$  recursively, and finally compute  $\pi$  by adding  $n$  to the  $i$ th block of  $\pi'$ .

**5.28. Example.** Let us compute the rank of  $\pi = \{\{1, 7\}, \{2, 4, 5\}, \{3, 8\}, \{6\}\}$  relative to the set  $SP(8, 4)$ . In the first stage of the recursion, removal of the largest element 8 from block 2 leaves the set partition  $\pi' = \{\{1, 7\}, \{2, 4, 5\}, \{3\}, \{6\}\}$ . Therefore,

$$r_{8,4}(\pi) = S(7, 3) + 2S(7, 4) + r_{7,4}(\pi') = 301 + 2 \cdot 350 + r_{7,4}(\pi').$$

(See Figure 2.21 for a table of Stirling numbers, which were calculated using the recursion for  $S(n, k)$ .) In the second stage, removing 7 from block 0 leaves the set partition  $\pi'' = \{\{1\}, \{2, 4, 5\}, \{3\}, \{6\}\}$ . Hence,

$$r_{7,4}(\pi') = S(6, 3) + 0S(6, 4) + r_{6,4}(\pi'') = 90 + r_{6,4}(\pi'').$$

In the third stage, removing the block  $\{6\}$  leaves the set partition  $\pi^{(3)} = \{\{1\}, \{2, 4, 5\}, \{3\}\}$ , and

$$r_{6,4}(\pi'') = r_{5,3}(\pi^{(3)}).$$

In the fourth stage, removing 5 from block 1 leaves the set partition  $\pi^{(4)} = \{\{1\}, \{2, 4\}, \{3\}\}$ , and

$$r_{5,3}(\pi^{(3)}) = S(4, 2) + 1S(4, 3) + r_{4,3}(\pi^{(4)}) = 7 + 6 + r_{4,3}(\pi^{(4)}).$$

In the fifth stage, removing 4 from block 1 leaves the set partition  $\pi^{(5)} = \{\{1\}, \{2\}, \{3\}\}$ , and

$$r_{4,3}(\pi^{(4)}) = S(3, 2) + 1S(3, 3) + r_{3,3}(\pi^{(5)}) = 3 + 1 + r_{3,3}(\pi^{(5)}).$$

But  $r_{3,3}(\pi^{(5)})$  is zero, since  $|SP(3, 3)| = 1$ . We deduce in sequence

$$r_{4,3}(\pi^{(4)}) = 4, \quad r_{6,4}(\pi'') = r_{5,3}(\pi^{(3)}) = 17, \quad r_{7,4}(\pi') = 107, \quad r_{8,4}(\pi) = 1108.$$

Next, let us compute  $u_{7,3}(111)$ . The input 111 weakly exceeds  $S(6, 2) = 31$ , so we must first compute  $p_{3,90}^{-1}(111 - 31) = (0, 80)$ . This means that 7 will go in block 0 of  $u_{6,3}(80)$ . Now  $80 \geq S(5, 2) = 15$ , so we compute  $p_{3,25}^{-1}(80 - 15) = (2, 15)$ . This means that 6 will go in block 2 of  $u_{5,3}(15)$ . Now  $15 \geq S(4, 2) = 7$ , so we compute  $p_{3,6}^{-1}(15 - 7) = (1, 2)$ . This means that 5 will go in block 1 of  $u_{4,3}(2)$ . Now  $2 < S(3, 2) = 3$ , so 4 is in a block by itself in  $u_{4,3}(2)$ . To find the remaining blocks, we compute  $u_{3,2}(2)$ . Now  $2 \geq S(2, 1) = 1$ , so we compute  $p_{2,1}^{-1}(2 - 1) = (1, 0)$ . This means that 3 goes in block 1 of  $u_{2,2}(0)$ . Evidently,  $u_{2,2}(0) = \{\{1\}, \{2\}\}$ . Using the preceding information to insert elements 3, 4, ..., 7, we conclude that

$$u_{7,3}(111) = \{\{1, 7\}, \{2, 3, 5\}, \{4, 6\}\}.$$

The ranking/unranking procedure given here lists the objects in  $SP(n, k)$  in the following order. Set partitions with  $n$  in its own block appear first. Next come the set partitions with  $n$  in block zero (i.e.,  $n$  is in the same block as 1); then come the set partitions with  $n$  in block one, etc. By applying the sum and product bijections in different orders, one can obtain different listings of the elements of  $SP(n, k)$ .

Let  $SP(n)$  be the set of all set partitions of  $n$ , so  $|SP(n)|$  is the  $n$ th Bell number. The preceding results lead to ranking and unranking algorithms for this collection, by applying the bijective sum rule to the disjoint union

$$SP(n) = SP(n, 1) \cup SP(n, 2) \cup \cdots \cup SP(n, n).$$

Another approach to ranking  $SP(n)$  is to use the recursion in 2.53. Details of this approach are left as an exercise.

## 5.10 Ranking Card Hands

We now apply the preceding ideas to the problem of ranking and unranking certain poker hands. We will use the bijective sum and product rules to transform the counting arguments from §1.13 into ranking and unranking bijections.

In this section, we define  $\text{Deck} = \text{Suits} \times \text{Values}$ , where

$$\text{Suits} = \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\};$$

$$\text{Values} = \{A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\}.$$

(A slightly different definition was used in §1.13.) The displayed orderings of the suits and values determine ranking and unranking bijections  $\text{Suits} \leftrightarrow \underline{4}$  and  $\text{Values} \leftrightarrow \underline{13}$ . For example,  $r(\diamondsuit) = 1 = r(2)$ ; 11 unranks to the value Q; and 3 unranks to the suit  $\spadesuit$ . Combining these maps with the map  $p_{4,13}$  from the bijective product rule, we obtain ranking and unranking bijections  $\text{Deck} \leftrightarrow \underline{52}$ . Since we are thinking of Deck as the product set  $\text{Suits} \times \text{Values}$ , the suit of a card is more significant than its value when ranking. With these conventions, the list of cards generated by unranking  $0, 1, \dots, 51$  in this order runs as follows:

$$A\clubsuit, 2\clubsuit, \dots, K\clubsuit, A\diamondsuit, 2\diamondsuit, \dots, K\diamondsuit, A\heartsuit, \dots, K\heartsuit, A\spadesuit, \dots, Q\spadesuit, K\spadesuit.$$

Naturally, all ranking and unranking results in the examples below depend on this chosen ordering of the deck.

Recall that a poker hand is a five-element subset of Deck. To rank or unrank such hands, we can use the bijections  $\text{Deck} \leftrightarrow \underline{52}$  to reduce to the problem of ranking and unranking five-element subsets of  $\underline{52}$ . This problem was solved in §5.6. More interesting ranking and unranking problems arise if we restrict attention to certain special kinds of hands, like a full house. We discuss some of these problems next; other examples are treated in the exercises.

**5.29. Example: Four-of-a-kind hands.** Let  $S$  be the set of all four-of-a-kind poker hands. Recall (§1.13) that we can build a hand  $H \in S$  (via the ordinary product rule) by picking one of the 13 values  $v \in \text{Values}$ , and then picking one of the 48 cards in  $\text{Deck} \sim (\text{Suits} \times \{v\})$ . The rank of the hand  $H$ , relative to this particular construction method for  $S$ , is

$$p_{13,48}(r(v), r'(c)),$$

where  $r : \text{Values} \rightarrow \underline{13}$  is the ranking function for card values, and  $r'$  is the ranking function

on Deck  $\sim (\text{Suits} \times \{v\})$  induced by the usual rank function on Deck. More precisely,  $r'(c)$  is the number of cards preceding  $c$  in the standard ordering of the deck after throwing out the four cards of value  $v$ . If  $c = (s_1, v_1)$ , one checks that  $r'(c) = 12r(s_1) + r(v_1) - \chi(v_1 > v)$ .

For example, let us rank the four-of-a-kind hand  $H = \{5\spadesuit, 8\heartsuit, 5\diamondsuit, 5\clubsuit, 5\heartsuit\}$ . Here,  $v = 5$  and  $c = 8\heartsuit = (\heartsuit, 8)$ . In the full deck,  $c$  has rank  $13 \cdot 2 + 7 = 33$ . But in the deck with the 5's removed,  $c$  has rank  $r'(c) = 12 \cdot 2 + 6 = 30$ . Accordingly,  $r(H) = p_{13,48}(4, 30) = 4 \cdot 48 + 30 = 222$ .

To illustrate unranking, let us compute  $u(600)$ . First,  $p_{13,48}^{-1}(600) = (12, 24)$ . It follows that  $v = u(12) = K$ . Next, unranking 24 relative to the deck with the four kings deleted gives us the card  $c = A\heartsuit$ . Therefore,  $u(600) = \{K\clubsuit, K\diamondsuit, K\heartsuit, K\spadesuit, A\heartsuit\}$ .

**5.30. Example: Full house hands.** Let  $S$  be the set of all full house hands. Recall (§1.13) that we can build a hand  $H \in S$  from the data  $(x, B, y, C)$ , where  $x \in \text{Values}$ ,  $B$  is a three-element subset of Suits,  $y \in \text{Values} \sim \{x\}$ , and  $C$  is a two-element subset of Suits. For example, the data

$$(x, B, y, C) = (J, \{\clubsuit, \diamondsuit, \spadesuit\}, 9, \{\clubsuit, \heartsuit\})$$

generate the full house hand  $H = \{J\clubsuit, J\diamondsuit, J\spadesuit, 9\clubsuit, 9\heartsuit\}$ . The rank of this hand is

$$p_{13,4,12,6}(r(x), r(B), r_x(y), r(C)),$$

where we use the same letter  $r$  to denote various ranking functions on the set of choices available at each stage. We write  $r_x(y)$  to emphasize that the ranking function for  $y \in \text{Values} \sim \{x\}$  depends on the value of the previous choice  $x$ . For the sample choice sequence considered above, we get

$$r(H) = p_{13,4,12,6}(10, 1, 8, 1) = 2880 + 72 + 48 + 1 = 3001.$$

(We are using the ranking functions for  $k$ -element subsets of Suits, which were discussed in §5.6.) Observe that the answer depends critically on the precise ordering of the choices we made in the counting argument. If we had chosen the data in the order  $(x, y, B, C)$ , for example, then we would obtain a different answer for  $r(H)$ .

To illustrate unranking, let us compute  $u(515)$ . First,

$$p_{13,4,12,6}^{-1}(515) = (1, 3, 1, 5).$$

Continuing to unrank,  $x = u(1) = 2$ ,  $B = u(3) = \{\diamondsuit, \heartsuit, \spadesuit\}$ ,  $y = u_x(1) = 3$  (since the value 2 has been deleted), and  $C = u(5) = \{\heartsuit, \spadesuit\}$ . So

$$u(515) = \{2\diamondsuit, 2\heartsuit, 2\spadesuit, 3\heartsuit, 3\spadesuit\}.$$

**5.31. Example: Two-pair hands.** Let  $S$  be the set of two-pair poker hands. This time, we build  $H \in S$  from data  $(B, C, D, z)$ , where  $B$  is a two-element subset of Values,  $C$  is a two-element subset of Suits,  $D$  is a two-element subset of Suits, and  $z = (s, v)$  is a card such that the value  $v$  is not in  $B$ . Using the bijective product rule, we have

$$r(H) = p_{78,6,6,44}(r(B), r(C), r(D), r_B(z)).$$

For example, let us find the rank of  $H = \{2\spadesuit, 2\clubsuit, 9\clubsuit, 9\heartsuit, K\diamondsuit\}$ , which arises from the data

$$(B, C, D, z) = (\{2, 9\}, \{\clubsuit, \spadesuit\}, \{\clubsuit, \heartsuit\}, (\diamondsuit, K)).$$

The ranking formula developed in §5.6 gives  $r(B) = \binom{1}{1} + \binom{8}{2} = 29$ ; similarly,  $r(C) = 3$  and  $r(D) = 1$ . After removing all 2's and 9's from the deck, the new rank of  $K\diamondsuit$  is  $r_B((\diamondsuit, K)) = 11 + 10 = 21$ . So, finally,

$$r(H) = 29 \cdot 6 \cdot 6 \cdot 44 + 3 \cdot 6 \cdot 44 + 1 \cdot 44 + 21 = 46,793.$$

**5.32. Example: Ordinary hands.** In §1.13, we built an “ordinary” poker hand  $H$  by choosing a five-element subset  $V(H)$  of Values that avoided one of the ten possible value sets for a straight, and then choosing a word in  $\text{Suits}^5$  not all of whose letters are equal (to avoid flushes). This argument showed that there are  $(C(13, 5) - 10) \cdot (4^5 - 4) = 1,302,540$  ordinary poker hands. How can we find a ranking algorithm for this collection of card hands?

Let  $Y$  be the set of all five-element subsets of Values, and let  $Z = \text{Suits}^5$ . We have already found ranking functions  $r_Y : Y \rightarrow \mathbf{C}(13, 5)$  and  $r_Z : Z \rightarrow \mathbf{4}^5$ . To take the prohibited conditions into account, let  $Y' = \{\{A, 2, 3, 4, 5\}, \{2, 3, 4, 5, 6\}, \dots\}$  be the set of ten objects in  $Y$  corresponding to straights, and let  $Z' = \{sssss : s \in \text{Suits}\}$  be the four objects in  $Z$  corresponding to flushes. We can get ranking functions  $r'_Y : Y \sim Y' \rightarrow \mathbf{C}(13, 5) - 10$  and  $r'_Z : Z \sim Z' \rightarrow \mathbf{4}^5 - 4$  by setting

$$r'_Y(C) = r_Y(C) - |\{C' \in Y' : r_Y(C') < r_Y(C)\}|.$$

This formula is practical since there are only ten possibilities for  $C'$ , and we can compute the ranks of these objects in advance. They are:

$$0, 5, 20, 55, 125, 251, 461, 791, 1278, 1286.$$

For example,  $r(\{3, 4, 5, 6, 7\}) = C(2, 1) + C(3, 2) + C(4, 3) + C(5, 4) + C(6, 5) = 20$  and  $r(\{10, J, Q, K, A\}) = C(0, 1) + C(9, 2) + C(10, 3) + C(11, 4) + C(12, 5) = 1278$ . Now, the rank function on  $Y \sim Y'$  can be computed via the formula

$$r'_Y(C) = \begin{cases} r_Y(C) - 1 & \text{if } 0 < r_Y(C) < 5; \\ r_Y(C) - 2 & \text{if } 5 < r_Y(C) < 20; \\ \dots & \dots \\ r_Y(C) - 9 & \text{if } 1278 < r_Y(C) < 1286. \end{cases}$$

Similarly, we can set

$$r'_Z(w) = r_Z(w) - |\{w' \in Z' : r_Z(w') < r_Z(w)\}|.$$

In this case, we precompute

$$\begin{aligned} r(\clubsuit\clubsuit\clubsuit\clubsuit\clubsuit) &= p_{4,4,4,4,4}(0, 0, 0, 0, 0) = 0; \\ r(\diamond\diamond\diamond\diamond\diamond) &= p_{4,4,4,4,4}(1, 1, 1, 1, 1) = 341; \\ r(\heartsuit\heartsuit\heartsuit\heartsuit\heartsuit) &= p_{4,4,4,4,4}(2, 2, 2, 2, 2) = 682; \\ r(\spadesuit\spadesuit\spadesuit\spadesuit\spadesuit) &= p_{4,4,4,4,4}(3, 3, 3, 3, 3) = 1023. \end{aligned}$$

Since these numbers form an arithmetic progression, we can write

$$r'_Z(w) = r_Z(w) - \lceil r_Z(w)/341 \rceil \quad (w \in Z \sim Z').$$

Finally, the overall ranking function for a hand  $H$  constructed from the pair  $(C, w)$  is given by  $r(H) = p_{1277, 1020}(r'_Y(C), r'_Z(w))$ . For example, let us compute the rank of the hand  $H = \{A\clubsuit, 4\clubsuit, 7\heartsuit, 9\clubsuit, 10\diamond\}$ . For this hand,  $C = \{A, 4, 7, 9, 10\}$  and  $w = \clubsuit\clubsuit\heartsuit\clubsuit\diamond$ . We calculate

$$r_Y(C) = \binom{0}{1} + \binom{3}{2} + \binom{6}{3} + \binom{8}{4} + \binom{9}{5} = 219; \quad r'_Y(C) = 219 - 5 = 214;$$

$$r_Z(w) = 2 \cdot 4^2 + 1 = 33; \quad r'_Z(w) = 33 - 1 = 32;$$

$$r(H) = p_{1277, 1020}(214, 32) = 214 \cdot 1020 + 32 = 218,312.$$

As another example, let us unrank 1,000,000. First,  $p_{1277,1020}^{-1}(10^6) = (980, 400)$ . The number 980 is between 791 and 1278 in the list of ranks of objects in  $Y'$ , so we recover  $r_Y(C) = 980 + 8 = 988$ . Unranking 988 produces the subset  $\{4, 5, 8, 9, 12\}$  of **13**, which translates into the value set  $\{5, 6, 9, 10, K\}$ . Next, since 400 lies between 341 and 682, we recover  $r_Z(w) = 402$ . Repeated division by 4 produces the base-4 number 12102, which translates to the value sequence  $z = \diamond \heartsuit \diamond \clubsuit \heartsuit$ . In conclusion,  $u(10^6) = \{5\diamond, 6\heartsuit, 9\diamond, 10\clubsuit, K\heartsuit\}$ .

This example shows that applications of the difference rule can be difficult to translate into ranking and unranking algorithms. Our success here depended on the fact that the sizes of the sets being subtracted were quite small, so that their effect on the ranking could be specified by a relatively brief case analysis.

**5.33. Remark.** In general, the orderings of special card hands obtained above do *not* necessarily arise by restricting the usual ordering of all five-card hands to the given subcollection. Rather, these orderings arise from the particular ordered sequence of choices used to generate these hands. Considerable cleverness may be required to find a ranking algorithm for generating a particular subcollection of card hands in lexicographic order.

## 5.11 Ranking Dyck Paths

Recall that a *Dyck path* of order  $n$  is a lattice path from  $(0, 0)$  to  $(n, n)$  that never goes below the line  $y = x$ . These objects are counted by the Catalan numbers  $C_n = \frac{1}{n+1} \binom{2n}{n}$ , which satisfy the recursion

$$C_n = \sum_{k=1}^n C_{k-1} C_{n-k} \quad (n > 0)$$

and initial condition  $C_0 = 1$  (see 2.33). Recall that the recursion classifies Dyck paths ending at  $(n, n)$  based on the first point  $(k, k)$  at which the path returns to the line  $y = x$  after leaving the origin (see Figure 2.10). If we use words  $w \in \{N, E\}^{2n}$  to encode Dyck paths, the first-return recursion corresponds to the factorization  $w = Nw_1Ew_2$ , where  $w_1$  encodes a Dyck path of order  $k - 1$ , and  $w_2$  encodes a Dyck path of order  $n - k$ .

By applying the bijective sum and product rules to the preceding recursion, we can obtain recursive ranking and unranking algorithms for Dyck paths. For each  $n \geq 0$ , let  $D_n$  be the set of words encoding Dyck paths of order  $n$ . Define ranking maps  $r_n : D_n \rightarrow \underline{\mathbf{C}}_n$  as follows. When  $n = 0$ ,  $r_0$  maps the empty word to the integer 0. For  $n > 0$ , suppose  $w \in D_n$  has first-return factorization  $w = Nw_1Ew_2$ , where  $w_1 \in C_{k-1}$  and  $w_2 \in C_{n-k}$  for some  $k$ . Recursively compute

$$r_n(w) = \sum_{j=1}^{k-1} C_{j-1} C_{n-j} + p_{C_{k-1}, C_{n-k}}(r_{k-1}(w_1), r_{n-k}(w_2)).$$

The unranking maps  $u_n : \underline{\mathbf{C}}_n \rightarrow D_n$  are defined as follows. First,  $u_0(0)$  is the empty word. Given  $n > 0$  and  $z \in \underline{\mathbf{C}}_n$ , find the unique integer  $k \leq n$  with

$$\sum_{j < k} C_{j-1} C_{n-j} \leq z < \sum_{j \leq k} C_{j-1} C_{n-j}.$$

Next, compute

$$(x, y) = p_{C_{k-1}, C_{n-k}}^{-1} \left( z - \sum_{j < k} C_{j-1} C_{n-j} \right).$$

Recursively determine the words  $w_1 = u_{k-1}(x)$  and  $w_2 = u_{n-k}(y)$ , and return the answer  $u_n(z) = Nw_1Ew_2$ .

**5.34. Example.** Let us compute the rank of the Dyck path  $w = \text{NNENNEEEENE}$ . The first-return factorization of  $w$  is  $w = Nw_1Ew_2$  where  $w_1 = \text{NENNEE}$  and  $w_2 = \text{NE}$ . Here,  $n = 5$ ,  $k - 1 = 3$ ,  $k = 4$ , and  $n - k = 1$ . The ranking formula gives

$$r_5(w) = C_0C_4 + C_1C_3 + C_2C_2 + p_{C_3, C_1}(r_3(w_1), r_1(w_2)).$$

Now  $r_1(w_2) = 0$  since  $w_2$  is the only Dyck path of order 1. As for  $r_3(w_1)$ , we proceed recursively. The required factorization of  $w_1$  is  $w_1 = Nw_3Ew_4$  where  $w_3$  is the empty word and  $w_4 = \text{NNEE}$ . At this stage of the recursive computation, we have  $n = 3$ ,  $k - 1 = 0$ ,  $k = 1$ , and  $n - k = 2$ . Therefore,

$$r_3(w_1) = p_{C_0, C_2}(r_0(w_3), r_2(w_4)).$$

Now  $r_0(w_3) = 0$ ; as for  $w_4$ , we have (writing  $\epsilon$  for the empty word)

$$r_2(w_4) = C_0C_1 + p_{C_1, C_0}(r_1(NE), r_0(\epsilon)) = 1 + p_{1,1}(0, 0) = 1.$$

Recall that the first few Catalan numbers are

$$C_0 = 1, C_1 = 1, C_2 = 2, C_3 = 5, C_4 = 14, C_5 = 42, C_6 = 132, C_7 = 429.$$

Working our way back through the recursive calculations, we find that  $r_3(w_1) = p_{1,2}(0, 1) = 1$  and

$$r_5(w) = 14 + 5 + 4 + p_{5,1}(1, 0) = 23 + 1 = 24.$$

**5.35. Remark.** The recursive ranking and unranking calculations can be simplified slightly by precomputing the ranks of Dyck paths of small order, which occur over and over again in the calculations. Observe that our ranking method for  $D_n$  lists the Dyck paths in the following order: first, all paths whose first return to  $y = x$  is at  $(1, 1)$ ; second, all paths whose first return is at  $(2, 2)$ ; and so on. Within each of these sublists, the ordering of paths is determined recursively (with the aid of the bijective product rule). Using these observations, we can quickly enumerate Dyck paths of order at most 3 in the order implied by the unranking algorithm. For  $n = 0$ , the list consists of just the empty word. For  $n = 1$ , we get: NE. For  $n = 2$ , we get: NENE, NNEE. For  $n = 3$ , we get:

$$\text{NENENE, NENNEE, NNEENE, NNENEE, NNNEEE.}$$

**5.36. Example.** Let us unrank 211 to obtain a Dyck path of order  $n = 7$ . From the recursion, we have

$$429 = C_7 = 1 \cdot 132 + 1 \cdot 42 + 2 \cdot 14 + 5 \cdot 5 + 14 \cdot 2 + 42 \cdot 1 + 132 \cdot 1.$$

The first step in unranking is to calculate the partial sums on the right side until we find the first one larger than the given input 211. We find that

$$C_0C_6 + C_1C_5 + C_2C_4 = 202 \leq 211 < 227 = C_0C_6 + C_1C_5 + C_2C_4 + C_3C_3.$$

Therefore  $k = 4$ ,  $k - 1 = 3 = n - k$ , and  $(x, y) = p_{5,5}^{-1}(211 - 202) = (1, 4)$ . Using the previous example, we have  $w_1 = u_3(x) = \text{NENNEE}$  and  $w_2 = u_3(y) = \text{NNNEEE}$ . It follows that

$$u_7(211) = \text{N NENNEE E NNNEEE.}$$

## 5.12 Ranking Trees

We know from §3.7 that there are  $n^{n-2}$  rooted trees on the vertex set  $\{1, 2, \dots, n\}$  rooted at vertex 1. Let  $B$  be the set of such trees; we seek ranking and unranking algorithms for  $B$ . One way to obtain these algorithms is to use the bijective proof of 3.47. In that proof, we described a bijection  $\phi' : B \rightarrow A$ , where  $A$  is the set of all functions  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  such that  $f(1) = 1$  and  $f(n) = n$ . Let  $C$  be the set of words of length  $n - 2$  in the alphabet  $\{0, 1, \dots, n - 1\}$ . The map  $\psi : A \rightarrow C$  such that  $\psi(f) = w_1 \cdots w_{n-2}$  with  $w_i = f(i+1) - 1$ , is evidently a bijection. Furthermore, the map  $p_{n,n,\dots,n}$  used in the bijective product rule gives a bijection from  $C$  to  $\underline{n}^{n-2}$ . Composing all these bijections, we get the desired ranking algorithm. Inverting the bijections gives an unranking algorithm.

**5.37. Example.** Consider the rooted tree  $T$  shown in Figure 3.9. In 3.49, we computed  $\phi'(T)$  to be the function  $g$  defined by

$$g : 1 \mapsto 1, 2 \mapsto 2, 3 \mapsto 2, 4 \mapsto 9, 5 \mapsto 9, 6 \mapsto 7, 7 \mapsto 6, 8 \mapsto 9, 9 \mapsto 9.$$

Here,  $\psi(g)$  is the word 1188658. Applying the map  $p_{9,9,\dots,9}$  (or equivalently, interpreting the given word as a number written in base 9), we find that  $r(T) = 649,349$ .

This application shows how valuable a bijective proof of a counting result can be. If we have a bijection from a complicated set of objects to a “nice” set of objects (such as functions or words), we can compose the bijection with standard ranking maps to obtain ranking and unranking algorithms for the complicated objects. In contrast, if a counting result is obtained by some intricate algebraic manipulation, it may not be so straightforward to extract an effective ranking mechanism.

## 5.13 Successors and Predecessors

Suppose  $S$  is a finite set of  $n$  objects, and we wish to list all the elements of  $S$  in a certain order. If we know an appropriate unranking bijection  $u : \underline{n} \rightarrow S$ , then we can generate the desired list by computing  $u(0), u(1), \dots, u(n-1)$  in succession. However, if the unranking map  $u$  is complicated, this method of listing  $S$  may not be very efficient.

In many applications, if we know the object  $z = u(i)$  that occupies a particular position on the list, it may be possible to compute the object that immediately precedes or follows  $z$  on the list, without ever explicitly computing  $i$  or applying the algorithm defining  $u$  to the inputs  $i - 1$  or  $i + 1$ . We call  $u(i - 1)$  the *predecessor* of  $z$  (relative to the listing determined by  $u$ ), and we call  $u(i + 1)$  the *successor* of  $z$ . Reversing the ordering of the elements of  $S$  interchanges predecessors and successors; so, in what follows, we need only consider successors.

The *successor problem* asks for an efficient algorithm for finding the successor of a given object  $z$  relative to a given ordering. We could solve this problem by ranking  $z$ , adding 1, and unranking, but we typically want more elegant solutions. If we can solve the successor problem, and if we know what the first object on the list is, then we will have a new method for listing all the elements of  $S$ . Namely, we start at the first object and then repeatedly invoke the successor algorithm until the last object is reached. In computer programming, one often uses this general strategy to loop through all elements of some set of combinatorial objects.



As a typical example, we develop a successor algorithm for listing the words in  $\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  in alphabetical order. This example includes as special cases the problem of listing all  $k$ -element subsets of an  $n$ -element set (which can be encoded as words in  $\mathcal{R}(0^{n-k}1^k)$ ) and the problem of listing all permutations of  $k$  letters (take each  $n_k = 1$ ). At the outset, we fix an ordered alphabet  $A = \{a_1 < \cdots < a_k\}$ . Our algorithm will consist of three functions, called “first,” “last,” and “next.” The “first” and “last” functions return the first and last words in  $\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  relative to the alphabetical ordering; explicitly, we have

$$\text{first}(n_1, \dots, n_k) = a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k};$$

$$\text{last}(n_1, \dots, n_k) = a_k^{n_k} a_{k-1}^{n_{k-1}} \cdots a_1^{n_1}.$$

In the case  $n_1 + \cdots + n_k = 0$ , both functions return the empty word.

The successor function “next” takes as input the integers  $n_1, \dots, n_k \geq 0$ , as well as a word  $z \in \mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$ . The output of “next” is either the successor  $z'$  of  $z$  in the alphabetical ordering, or a special flag (called “done”) indicating that  $z$  was the last word on the list. The operation of “next” is based on the observation that the alphabetical list of words consists of all words starting with  $a_1$  (if  $n_1 > 0$ ), then all words starting with  $a_2$  (if  $n_2 > 0$ ), and so on. Within each of these sublists, the words are ordered in the same way based on their second letters, and so on recursively. So we can define  $\text{next}(n_1, \dots, n_k, z)$  using the following recursive algorithm.

- Base Case: If  $n_1 + \cdots + n_k \leq 0$  or if  $z = \text{last}(n_1, \dots, n_k)$ , return “done.”
- Recursion: Suppose  $z = a_i z'$ . Recursively compute  $w' = \text{next}(n_1, \dots, n_i - 1, \dots, n_k, z')$ . If  $w'$  is not the special “done” flag, return  $a_i w'$  as the answer. Otherwise, find the smallest index  $j > i$  with  $n_j > 0$ . If no such index exists, return “done”; else return the concatenation of  $a_j$  and  $\text{first}(n_1, \dots, n_j - 1, \dots, n_k)$ .

**5.38. Example.** Consider  $z = \text{bdcacbc} \in \mathcal{R}(a^1 b^2 c^3 d^1)$ . To compute the successor of  $z$ , we are first directed to find the successor of  $z' = \text{dcacbc} \in \mathcal{R}(a^1 b^1 c^3 d^1)$ . Continuing recursively, we are led to consider the words  $\text{cacbc}$ , then  $\text{accbc}$ , then  $\text{ccbc}$ . But  $\text{ccbc}$  is the last word in  $\mathcal{R}(a^0 b^1 c^2 d^0)$ , so  $\text{next}(0, 1, 2, 0, \text{ccbc})$  returns “done.” Returning to the calculation of  $\text{next}(1, 1, 2, 0, \text{accbc})$ , we seek the next available letter after ‘a’, which is ‘b’. We concatenate ‘b’ and  $\text{first}(1, 0, 2, 0) = \text{acc}$  to obtain  $\text{next}(1, 1, 2, 0, \text{accbc}) = \text{bacc}$ . Then  $\text{next}(1, 1, 3, 0, \text{cacbc}) = \text{cbacc}$ . Continuing similarly, we eventually obtain the final output “bdcbacc” as the successor of  $z$ .

If we apply the “next” function to this new word, we strip off initial letters one at a time until we reach the suffix “cc,” which is the last word in its class. So, to compute  $\text{next}(1, 0, 2, 0, \text{acc})$ , we must find the next *available* letter after ‘a’, namely ‘c’, and append to this the word  $\text{first}(1, 0, 1, 0) = \text{ac}$ . Thus,  $\text{next}(1, 0, 2, 0, \text{acc}) = \text{cac}$ , and working back up the recursive calls leads to a final answer of “bdcbcac.” This example shows why we need to remember the values  $n_1, \dots, n_k$  in each recursive call to “next.”

**5.39. Example.** If we use the given method to list the permutations of  $\{1, 2, 3, 4\}$ , we obtain:

1234, 1243, 1324, 1342, 1423, 1432, 2134, 2143, ..., 4321.

**5.40. Example.** Using the encoding of subsets as binary words (see 1.38), we can list all 2-element subsets of  $\{1, 2, 3, 4, 5\}$  by running through the words in  $\mathcal{R}(1^2 0^3)$ . (For convenience, we choose the alphabet ordering  $1 < 0$  here.) The words are:

11000, 10100, 10010, 10001, 01100, 01010, 01001, 00110, 00101, 00011.

The associated subsets are:

$$\{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 4\}, \{2, 5\}, \{3, 4\}, \{3, 5\}, \{4, 5\}.$$

In general, the method used here lists  $k$ -element subsets of  $\{1, 2, \dots, n\}$  in lexicographic order. Using the ordering of the letters  $0 < 1$  would have produced the reversal of the list displayed above. In contrast, the ranking method discussed in §5.6 lists the subsets according to a different ordering, in which all subsets not containing  $n$  are listed first, followed by all subsets that do contain  $n$ , and so on recursively. The latter method produces the following list of subsets:

$$\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 4\}, \{2, 4\}, \{3, 4\}, \{1, 5\}, \{2, 5\}, \{3, 5\}, \{4, 5\}.$$

**5.41. Example: Successor Algorithm for Dyck Paths.** Let us design a successor algorithm for generating Dyck paths of order  $n$ , based on the “first-return” recursion (§2.7). As above, we will use three routines called “first,” “last,” and “next.” The routine  $\text{first}(n)$  returns the path  $(NE)^n$ , which returns to the diagonal as early as possible, while the routine  $\text{last}(n)$  returns the path  $N^nE^n$ , which returns to the diagonal as late as possible. If  $n \geq 0$  and  $w$  encodes a Dyck path of order  $n$ , then  $\text{next}(n, w)$  will return the next Dyck path in the chosen ordering, or “done” if  $w$  is the last path. This routine works as follows:

- Base Case: If  $n \leq 1$  or  $w = \text{last}(n)$ , then return “done.”
- Recursion: Find the first-return factorization  $w = Nw_1Ew_2$  of  $w$ , where  $w_1 \in D_{k-1}$  and  $w_2 \in D_{n-k}$  for some  $k$  between 1 and  $n$ . Now consider subcases.
  - Calculate  $w'_2 = \text{next}(n - k, w_2)$ . If this path exists, return  $Nw_1Ew'_2$  as the answer.
  - Otherwise, find  $w'_1 = \text{next}(k - 1, w_1)$ . If this path exists, return  $Nw'_1E\text{first}(n - k)$  as the answer.
  - Otherwise, increase  $k$  by 1. If the new  $k$  is  $\leq n$ , return  $N\text{first}(k - 1)E\text{first}(n - k)$  as the answer.
  - Otherwise, return “done.”

For example, let us compute  $\text{next}(7, N\text{NENNEE}E\text{NNNEEEE})$ . The given input  $w$  factorizes as  $w = Nw_1Ew_2$  where  $w_1 = \text{NENNEE}$  and  $w_2 = \text{NNNEEEE}$ . Here,  $k = 4$ ,  $k - 1 = 3$ , and  $n - k = 3$ . By inspection,  $w_2 = \text{last}(3)$ , so we proceed to the next subcase. We are directed to compute  $w'_1 = \text{next}(3, w_1)$ . Here,  $w_1$  factors as  $w_1 = Nw_3Ew_4$ , where  $w_3$  is the empty word and  $w_4 = \text{NNEE}$ . Again,  $w_4$  is the last Dyck path of order 2, and this time  $w_3$  is also the last Dyck path of order 0. So we increase  $k$  by 1, and return

$$w'_1 = N\text{first}(1)E\text{first}(1) = N\text{NE}E\text{NE}.$$

Using this result in the original calculation, we obtain

$$\text{next}(w) = Nw'_1E\text{first}(3) = N\text{NNEENE}E\text{NENENE}.$$

## 5.14 Random Selection

We now briefly revisit the problem of random selection of combinatorial objects. Suppose  $S$  is a set of  $n$  objects, and we wish to randomly select an element of  $S$ . If we have an

unranking function  $u : \underline{n} \rightarrow S$ , we can select the object by generating a random integer in  $\underline{n}$  and unranking it. However, it may be impractical to generate such an integer if  $n$  is very large. If the objects in  $S$  are generated by a recursive process, we can often get around this problem by making several random choices that are used to build an object in  $S$  in stages.

**5.42. Example: Subsets.** As a typical example, consider the question of randomly choosing a  $k$ -element subset of  $A = \{x_1, \dots, x_n\}$ . Earlier (§5.6), we developed ranking and unranking algorithms for this problem based on the recursion

$$C(n, k) = C(n - 1, k) + C(n - 1, k - 1). \quad (5.2)$$

A slight adjustment of the methods used before will lead to an efficient solution for the random selection problem.

Recall that the term  $C(n - 1, k)$  in the recursion counts the  $k$ -element subsets of  $A$  that do not contain  $x_n$ , while the term  $C(n - 1, k - 1)$  counts subsets that do contain  $x_n$ . If we choose a random  $k$ -element subset of  $A$ , the probability that it will not contain  $x_n$  is  $C(n - 1, k)/C(n, k) = (n - k)/n$ , and the probability that it will contain  $x_n$  is  $C(n - 1, k - 1)/C(n, k) = k/n$ . This suggests the following recursively defined random selection procedure. To choose a  $k$ -element subset  $A$  of  $\{x_1, \dots, x_n\}$ , use a random number generator to obtain a real number  $r \in [0, 1]$ . If  $r \leq k/n$ , declare that  $x_n$  belongs to  $A$ , and recursively select a random  $(k - 1)$ -element subset of  $\{x_1, \dots, x_{n-1}\}$  by the same method. If  $r > k/n$ , declare that  $x_n$  does not belong to  $A$ , and recursively select a random  $k$ -element subset of  $\{x_1, \dots, x_{n-1}\}$ . The base cases occur when  $k = 0$  or  $k = n$ , in which case there is only one possible subset to select.

**5.43. Example: Set Partitions.** A similar method can be used to randomly select a  $k$ -element set partition of an  $n$ -element set  $A = \{x_1, \dots, x_n\}$ . Recall that these objects are counted by the Stirling numbers  $S(n, k)$ , which satisfy the recursion

$$S(n, k) = S(n - 1, k - 1) + kS(n - 1, k) \quad (n, k > 0).$$

Recall that the first term on the right counts set partitions with  $x_n$  in a block by itself, while the second term counts set partitions in which  $x_n$  occurs in a block with other elements. The base cases of the selection procedure occur when  $k = 0$  or  $k = n$ ; here, there is at most one possible object to select. For the main case, assume  $n > 0$  and  $0 < k < n$ . Choose a random real number  $r \in [0, 1]$ . Consider the quantity  $r_0 = S(n - 1, k - 1)/S(n, k)$ , which can be computed (at least approximately) using the Stirling recursion. Note that  $r_0$  is the probability that a random set partition of  $A$  into  $k$  blocks will have  $x_n$  in a block by itself. If  $r \leq r_0$ , recursively select a set partition of  $\{x_1, \dots, x_{n-1}\}$  into  $k - 1$  blocks, and append the block  $\{x_n\}$  to this set partition to obtain the answer. In the alternative case  $r > r_0$ , recursively select a set partition  $\{B_1, \dots, B_k\}$  of  $\{x_1, \dots, x_{n-1}\}$  into  $k$  blocks. Now, choose a random integer  $i$  in the range  $\{1, \dots, k\}$ , and insert  $n$  into block  $B_i$  to obtain the answer. Such an integer  $i$  can be found, for example, by choosing another random real number  $s \in [0, 1]$ , multiplying by  $k$ , and rounding up to the nearest integer.

**5.44. Example: Permutations.** As a final example, consider the problem of randomly generating a permutation of  $\{1, 2, \dots, n\}$ . We can convert the standard counting argument (based on the product rule) into a random selection procedure. However, some care is required, since the available choices at each stage depend on what choices were made in previous stages.

Recall that one method for building a permutation  $w = w_1 w_2 \cdots w_n$  of  $\{1, 2, \dots, n\}$  is to choose  $w_1$  to be any letter (in  $n$  ways), then choosing  $w_2$  to be any letter other than  $w_1$  (in  $n - 1$  ways), then choosing  $w_3$  to be any letter other than  $w_1$  or  $w_2$  (in  $n - 2$  ways),

and so on. The associated random generation algorithm would operate as follows. Generate random integers  $i_1 \in \{1, 2, \dots, n\}$ ,  $i_2 \in \{1, 2, \dots, n-1\}$ , ...,  $i_n \in \{1\}$ . Define  $w_1 = i_1$ . Define  $w_2$  to be the  $i_2$ th smallest element in the set  $\{1, 2, \dots, n\} \sim \{w_1\}$ . In general, define  $w_j$  to be the  $i_j$ th smallest element in the set  $\{1, 2, \dots, n\} \sim \{w_1, \dots, w_{j-1}\}$ .

This computation can become rather messy, since we must repeatedly scan through the remaining letters to determine the  $i_j$ th smallest one. Furthermore, the method is really no different from unranking a random integer between 0 and  $n! - 1$ . An alternative recursive generation method proceeds as follows. If  $n = 1$ , return  $w = 1$  as the answer. If  $n > 1$ , first recursively generate a random permutation  $w' = w'_1 \cdots w'_{n-1}$  of  $\{1, 2, \dots, n-1\}$ . Next, generate a random integer  $w_n \in \{1, 2, \dots, n\}$ . Now make a single scan through the previously chosen letters  $w'_i$ , and let  $w_i = w'_i$  if  $w'_i < w_n$ ,  $w_i = w'_i + 1$  if  $w'_i \geq w_n$ . This determines the final answer  $w = w_1 w_2 \cdots w_n$ . We let the reader check that every permutation is equally likely to be generated when using this method.

## Summary

- *Definitions.* Let  $S$  be a set of  $n$  objects. A *ranking map* for  $S$  is a bijection  $r : S \rightarrow \underline{n}$ , where  $\underline{n} = \{0, 1, 2, \dots, n-1\}$ . An *unranking map* for  $S$  is a bijection  $u : \underline{n} \rightarrow S$ . Given a particular total ordering of the elements of  $S$ , a *successor map* for  $S$  is a function that maps each  $z \in S$  to the element immediately following  $z$  in the ordering (if any).
- *Bijective Sum Rule.* Let  $S_1, \dots, S_k$  be disjoint finite sets with union  $S$ . Given bijections  $f_i : S_i \rightarrow \underline{n_i}$ , there is a bijection  $f = \sum_i f_i : S \rightarrow \underline{n}$  (where  $n = n_1 + \cdots + n_k$ ) given by

$$f(x) = \sum_{j < i} n_j + f_i(x) \quad (x \in S_i).$$

The map  $f$  depends on the given ordering of the  $f_i$ 's. To compute  $f^{-1}(z)$ , find the unique index  $i$  such that  $\sum_{j < i} n_j \leq z < \sum_{j \leq i} n_j$ , and let  $f^{-1}(z) = f_i^{-1}(z - \sum_{j < i} n_j)$ .

- *Bijective Product Rule.* Given positive integers  $n_1, \dots, n_k$  with product  $n$ , there is a bijection  $p = p_{n_1, n_2, \dots, n_k} : \underline{n_1} \times \underline{n_2} \times \cdots \times \underline{n_k} \rightarrow \underline{n}$  given by

$$p(c_1, c_2, \dots, c_k) = c_1 n_2 \cdots n_k + c_2 n_3 \cdots n_k + \cdots + c_{k-1} n_k + c_k \quad (0 \leq c_i < n_i).$$

To compute  $p^{-1}(z)$ , let  $q_k$  and  $r_k$  be the quotient and remainder when  $z$  is divided by  $n_k$ . Then let  $q_{k-1}$  and  $r_{k-1}$  be the quotient and remainder when  $q_k$  is divided by  $n_{k-1}$ . Continue similarly; then  $p^{-1}(z) = (r_1, r_2, \dots, r_k)$ . If all  $n_i$ 's equal the same integer  $b$ ,  $p^{-1}(z)$  is the base- $b$  expansion of  $z$ .

- *Ranking  $k$ -Permutations.* Suppose we are ranking  $k$ -permutations  $w = w_1 w_2 \cdots w_k$  of an ordered alphabet  $A = (x_0, x_1, \dots, x_{n-1})$ . To find  $r(w)$ , first compute  $(j_1, \dots, j_k)$  by letting  $j_i$  be the number of letters preceding  $w_i$  in the given ordering of  $A$  that are different from  $w_1, \dots, w_{i-1}$ . Then calculate  $r(w) = p_{n, n-1, \dots, n-k+1}(j_1, \dots, j_k)$ . To unrank  $z$ , apply  $p^{-1}$  to recover  $(j_1, \dots, j_k)$ , and then recover  $w_1, \dots, w_k$  from left to right by letting  $w_i$  be the  $(j_i + 1)$ th smallest letter in  $A \sim \{w_1, \dots, w_{i-1}\}$ .
- *Rank Formula for Subsets.* Suppose we are ranking  $k$ -element subsets of an ordered alphabet  $A = (x_0, x_1, \dots, x_{n-1})$ . If  $B = \{x_{i_1} < x_{i_2} < \cdots < x_{i_k}\}$ , then we can take  $r(B) = \sum_{j=1}^k \binom{i_j}{j}$ . We can unrank a given integer  $z$  by a greedy strategy that recovers

$i_k, \dots, i_1$  (in this order) by choosing the largest possible value that will not cause the partial sum so far to exceed  $z$ . This ranking method leads to a listing of  $k$ -element subsets in which all subsets containing  $x_{n-1}$  appear after all subsets not containing  $x_{n-1}$ ; each sublist is ordered in the same way relative to  $x_{n-2}$ , etc.

- *Rank Formula for Anagrams.* The following recursive formula can be used to rank words  $w \in \mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  in alphabetical order: if  $w = a_i w'$ , then

$$r(w) = \sum_{j < i} C(n_1 + \cdots + n_k - 1; n_1, \dots, n_j - 1, \dots, n_k) + r(w').$$

To unrank a given integer  $z$ , choose  $i$  as large as possible so that the sum in the previous formula does not exceed  $z$ ; subtract the sum for this choice of  $i$  from  $z$ ; unrank the result recursively; and prepend the letter  $a_i$  to obtain the final answer.

- *Successor Algorithm for Anagrams.* Consider all words in  $\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  in alphabetical order. To find the word immediately following  $w$ , first write  $w = a_i w'$ . If  $w'$  is not the last word in its rearrangement class, recursively compute its successor (say  $z'$ ), and return  $a_i z'$  as the successor of  $w$ . Otherwise, find the first  $j > i$  with  $n_j > 0$ , and return  $a_j b'$  as the successor of  $w$ , where  $b' = a_1^{n_1} \cdots a_j^{n_j-1} \cdots a_k^{n_k}$ .
- *Random Selection Algorithms.* Suppose we want to randomly select an object from a given set  $S$ . If an unranking map  $u : \underline{n} \rightarrow S$  is available, we can generate a random integer  $z \in \underline{n}$  and return  $u(z)$ . Alternatively, if the objects in  $S$  can be built up in stages, we can make a random choice at each stage to decide how to build the object. For instance, to build a random  $k$ -subset  $B$  of  $\{1, 2, \dots, n\}$ , we can include  $n$  in  $B$  with probability  $k/n$ , and then choose the remaining elements of  $B$  recursively.

## Exercises

**5.45.** Suppose  $f : \{a, b, c\} \rightarrow \underline{3}$  and  $g : \{d, e\} \rightarrow \underline{2}$  are defined by  $f(a) = 1$ ,  $f(b) = 2$ ,  $f(c) = 0$ ,  $g(d) = 1$ ,  $g(e) = 0$ . Compute the bijections  $f + g$  and  $g + f$ .

**5.46.** Compute (a)  $p_{7,5}(4, 3)$ ; (b)  $p_{7,5}(3, 4)$ ; (c)  $p_{5,7}(4, 3)$ ; (d)  $p_{5,7}(3, 4)$ ; (e)  $p_{7,5}^{-1}(22)$ ; (f)  $p_{5,7}^{-1}(22)$ .

**5.47.** Find (a)  $p_{2,2,2,2}(0, 1, 1, 0, 1)$ ; (b)  $p_{2,2,2,2}^{-1}(29)$ ; (c)  $p_{7,7,7}(3, 0, 6)$ ; (d)  $p_{7,7,7}^{-1}(306)$ ; (e)  $p_{10,10,10}^{-1}(306)$ .

**5.48.** Compute: (a)  $p_{5,4,3,2,1}(3, 3, 0, 1, 0)$ ; (b)  $p_{5,4,3,2,1}^{-1}(111)$ ; (c)  $p_{3,6,2,6}(2, 5, 0, 4)$ ; (d)  $p_{3,6,2,6}^{-1}(150)$ ; (e)  $p_{6,2,6,3}^{-1}(150)$ ; (f)  $p_{6,6,3,2}^{-1}(150)$ .

**5.49.** Consider the product set  $X = \underline{3} \times \underline{4}$ . (a) View  $X$  as the disjoint union of the sets  $X_i = \{i\} \times \underline{4}$ , for  $i = 0, 1, 2$ . Let  $f_i : X_i \rightarrow \underline{4}$  be the bijection  $f_i(i, y) = y$ . Compute the bijections  $f_0 + f_1 + f_2$  and  $f_2 + f_1 + f_0$ , which map  $X$  to  $\underline{12}$ . (b) View  $X$  as the disjoint union of the sets  $X^{(j)} = \underline{3} \times \{j\}$ , for  $j = 0, 1, 2, 3$ . Let  $g_j : X^{(j)} \rightarrow \underline{3}$  be the bijection  $g_j(x, j) = x$ . Compute the bijection  $g_0 + g_1 + g_2 + g_3 : X \rightarrow \underline{12}$ . (c) Compute the bijection  $p_{3,4} : X \rightarrow \underline{12}$ . Is this one of the maps found in (a) or (b)? (d) Let  $t : X \rightarrow \underline{4} \times \underline{3}$  be the bijection  $t(i, j) = (j, i)$ . Compute the bijection  $p_{4,3} \circ t : X \rightarrow \underline{12}$ . Is this one of the maps found in (a) or (b)?

**5.50.** Rank the following four-letter words: (a) alto; (b) zone; (c) rank; (d) four; (e) word.

**5.51.** Unrank the following numbers in  $26^4$  to obtain four-letter words: (a) 115, 287; (b) 396, 588; (c) 392, 581; (d) 338, 902; (e) 275, 497.

**5.52.** (a) Rank the six-letter word “unrank.” (b) Unrank 199,247,301 to get a 6-letter word. (c) What happens if we unrank 199,247,301 to get a  $k$ -letter word where  $k > 6$ ?

**5.53.** A fraternity name consists of either two or three capital Greek letters. Recall that there are 24 letters in the Greek alphabet, ordered as follows:

$$\text{AB}\Gamma\Delta\text{E}\text{Z}\text{H}\Theta\text{I}\text{K}\Lambda\text{M}\text{N}\Xi\text{O}\Pi\rho\Sigma\Upsilon\Phi\text{X}\Psi\Omega.$$

Assume an ordering of fraternity names consisting of all two-letter names in alphabetical order, followed by all three-letter names in alphabetical order. Compute the rank of (a)  $\Phi\text{BK}$ ; (b)  $\Delta\Delta$ ; (c)  $\Delta\Delta\Delta$ ; (d)  $\text{AX}\Omega$ . Now, unrank: (e) 144; (f) 1440; (g) 13931.

**5.54.** Repeat (a)–(g) in 5.53, assuming the names are ordered so that all three-letter names precede all two-letter names, with names of each length in alphabetical order.

**5.55.** Repeat (a)–(g) in 5.53, assuming the names are ordered in alphabetical order (so that, for example,  $\Delta\Delta$  is immediately preceded by  $\Delta\Gamma\Omega$  and immediately followed by  $\Delta\Delta\Delta$ ).

**5.56.** Consider the set of four-digit even numbers (no leading zeroes allowed) that do not contain the digit 6. (a) Use the product rule to count this set. (b) Find a ranking bijection that will list these numbers in increasing numerical order. (c) Use (b) to rank 1234, 2500, and 9708. (d) Now unrank 1234, 2501, and 666.

**5.57.** Consider five-letter palindromes, ranked in alphabetical order. (a) Rank the palindromes LEVEL and MADAM. (b) Unrank 1581 and 12,662. (c) Find the first and last palindromes in the ranking that are real English words.

**5.58.** A Virginia license plate consists of three uppercase letters followed by four digits. For arcane bureaucratic reasons, license plate 0 is ZZZ-9999, followed by ZZZ-9998, ..., ZZZ-0000, ZZY-9999, etc. Use this system to rank the license plates: (a) ZCF-2073; (b) JXB-2007; (c) ABC-1234. Now unrank: (d) 7,777,777; (e) 123,456,789.

**5.59.** Repeat the previous exercise assuming a new ordering, honoring the 400th anniversary of Jamestown, where license plate 0 is JAM-1607, and license plates count forward in lexicographic order (“wrapping around” from ZZZ-9999 to AAA-0000).

**5.60.** Let  $A = \{a, b, c, d, e, f\}$ . (a) Compute the ranks of  $\text{bfdc}$  and  $\text{fdac}$  among all 4-permutations of  $A$ . (b) Unrank 232 to get a 4-permutation of  $A$ . (c) Compute the rank of  $\text{ecafdb}$  among all permutations of  $A$ . (d) Unrank 583 to get a permutation of  $A$ .

**5.61.** (a) Compute the rank of 42153 among all permutations of  $\{1, 2, \dots, 5\}$ . (b) Unrank 46 to obtain a permutation of  $\{1, 2, \dots, 5\}$ .

**5.62.** (a) Compute the rank of 36281745 among all permutations of  $\{1, 2, \dots, 8\}$ . (b) Unrank 23,419 to obtain a permutation of  $\{1, 2, \dots, 8\}$ .

**5.63.** Let  $A = \{a, b, c, d, e, f, g, h\}$ . (a) Use the ranking formula for  $S_4(A)$  in §5.6 to rank the subsets  $\{a, c, e, g\}$ ,  $\{b, c, d, h\}$ , and  $\{d, e, f, h\}$ . (b) Unrank 30, 40, and 50 to obtain 4-element subsets of  $A$ .

**5.64.** (a) Devise a ranking algorithm for  $k$ -element subsets of an  $n$ -element alphabet based on the recursion  $C(n, k) = C(n-1, k-1) + C(n-1, k)$ , which differs from the recursion in §5.6 due to the reversal of the order of terms on the right side. (b) Describe informally the order in which the ranking algorithm in (a) will produce the  $k$ -element subsets. (c) Answer the ranking and unranking questions in the previous exercise using this new ranking algorithm.

**5.65.** (a) Find the ranks of bbccacba and cabcabbc in the set  $\mathcal{R}(a^2b^3c^3)$ , ordered alphabetically. (b) Unrank 206 and 497 to get anagrams in  $\mathcal{R}(a^2b^3c^3)$ .

**5.66.** (a) Compute the rank of MISSISSIPPI among the set of all anagrams in  $\mathcal{R}(I^4MP^2S^4)$  (listed alphabetically). (b) Which anagram in this set has rank 33,333?

**5.67.** (a) Use the rank functions  $r_{n,k}$  in §5.8 to rank the integer partitions  $(3, 3, 3)$ ,  $(5, 2, 2)$ , and  $(4, 3, 2, 1)$ . (b) Compute  $u_{12,3}(6)$ ,  $u_{15,4}(22)$ , and  $u_{20,6}(47)$ .

**5.68.** Use the rank function  $r_{8,4}$  from §5.8 to list all integer partitions of 8 into 4 parts.

**5.69.** Enumerate all the integer partitions of 7, following the method used in 5.27.

**5.70.** Use the rank functions from §5.9 to list all set partitions of  $\{1, 2, 3, 4, 5\}$  into 3 blocks.

**5.71.** (a) Use the algorithms in §5.9 to rank the following set partitions relative to the set  $SP(n, k)$ :  $\{\{1, 3\}, \{2, 4, 5\}\}$ ;  $\{\{1, 5, 7\}, \{2\}, \{3, 4, 8\}, \{6\}\}$ . (b) Unrank 247 to obtain a set partition in  $SP(7, 4)$ . (c) Unrank 1492 to obtain a set partition in  $SP(8, 4)$ .

**5.72.** (a) Rank the four-of-a-kind hand  $\{3\clubsuit, 8\heartsuit, 8\diamondsuit, 8\spadesuit, 8\clubsuit\}$  (see 5.29). (b) Unrank 264 to get a four-of-a-kind hand.

**5.73.** (a) Rank the full house hand  $\{3\clubsuit, 3\heartsuit, 3\diamondsuit, 9\spadesuit, 9\diamondsuit\}$  (see 5.30). (b) Unrank 3082 to get a full house hand.

**5.74.** (a) Rank the full house hand  $\{A\spadesuit, A\heartsuit, A\diamondsuit, K\heartsuit, K\clubsuit\}$  (see 5.30). (b) Unrank 483 to get a full house hand.

**5.75.** (a) Rank the two-pair hand  $\{A\diamondsuit, A\heartsuit, 7\spadesuit, Q\diamondsuit, Q\spadesuit\}$  (see 5.31). (b) Unrank 71,031 to get a two-pair hand.

**5.76.** (a) Rank the hand  $\{5\clubsuit, 7\diamondsuit, 8\spadesuit, 10\heartsuit, J\heartsuit\}$  among all possible poker hands. (b) Rank the same hand among all “ordinary” poker hands (see 5.32). (c) Unrank 1,159,403 to get one of the  $\binom{52}{5}$  possible poker hands. (d) Unrank 1,159,403 to get an ordinary poker hand.

**5.77.** Use the ranking algorithm in §5.11 to list all Dyck paths of order (a) 4; (b) 5.

**5.78.** (a) Rank the Dyck path NNNENEENNEENNEEE (see §5.11). (b) Unrank 52 to get a Dyck path of order 6. (c) Unrank 335 to get a Dyck path of order 7.

**5.79.** (a) Use the method of §5.12 to find the rank of the rooted tree

$$T = \{(1, 1), (2, 1), (3, 1), (4, 1), (10, 1), (7, 1), (6, 7), (5, 6), (8, 5), (11, 6), (9, 11)\}.$$

(b) Unrank 1,609,765 to obtain a rooted tree on 9 vertices rooted at vertex 1.

**5.80.** Use the algorithm in §5.13 to find the successor of each word in the appropriate set of anagrams: (a) ccbabdc; (b) 3641275; (c) 01101011; (d) 33212312; (e) UKULELE.

**5.81.** Write an algorithm to find the predecessor of a given word  $w \in \mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$  in the alphabetical ordering of anagrams. Use this to find the predecessor of each word in the previous exercise.

**5.82.** (a) Use the successor algorithm in §5.13 to find the first four successors of the Dyck path NNNENEENNEENNENEE. (b) Write a predecessor algorithm for Dyck paths. (c) Use (b) to compute the first four predecessors of the Dyck path in (a).

**5.83.** Give careful proofs of the bijective sum rules (5.1 and 5.2).

**5.84.** Prove that for all  $a \in \mathbb{Z}$  and all nonzero  $b \in \mathbb{Z}$ , there exist unique  $q, r \in \mathbb{Z}$  with  $a = bq + r$  and  $0 \leq r < |b|$ . Describe an algorithm for computing  $q$  and  $r$  given  $a$  and  $b$ .

**5.85.** Prove that for all  $a \in \mathbb{Z}$  and all nonzero  $b \in \mathbb{Z}$ , there exist unique  $q, r \in \mathbb{Z}$  with  $a = bq + r$  and  $-|b|/2 < r \leq |b|/2$ . Describe an algorithm for computing  $q$  and  $r$  given  $a$  and  $b$ .

**5.86.** Suppose  $0 \neq b \in \mathbb{Z}$  and  $S \subseteq \mathbb{Z}$ . Find a necessary and sufficient condition on  $S$  that will make the following statement true: for all  $a \in \mathbb{Z}$ , there exist unique  $q, r \in \mathbb{Z}$  with  $a = bq + r$  and  $r \in S$ . How could one find  $q$  and  $r$  given  $a$ ,  $b$ , and  $S$ ?

**5.87. Division Algorithm for Polynomials.** (a) Suppose  $F$  is a field,  $g \in F[x]$  is a nonzero polynomial, and  $f \in F[x]$  is any polynomial. Prove there exist unique polynomials  $q, r \in F[x]$  with  $f = qg + r$  and either  $r = 0$  or  $\deg(r) < \deg(g)$ . Describe an algorithm for computing  $q$  and  $r$  given  $f$  and  $g$ . (b) Show that (a) can fail if  $F$  is a commutative ring that is not a field. (c) Is (a) true for all commutative rings  $F$  if we assume  $g$  is monic?

**5.88.** Verify 5.14.

**5.89.** Given  $a, b \in \mathbb{Z}$  with  $b > 0$ , let  $a \bmod b$  be the unique remainder  $r \in \underline{b}$  such that  $a = bq + r$  for some  $q \in \mathbb{Z}$ . (a) Given  $s, t > 0$ , consider the map  $f : \underline{st} \rightarrow \underline{s} \times \underline{t}$  given by  $f(x) = (x \bmod s, x \bmod t)$  for  $x \in \underline{st}$ . Prove that  $f$  is a bijection iff  $\gcd(s, t) = 1$ . (b) Generalize (a) to maps from  $\underline{s_1 s_2 \cdots s_k}$  to  $\underline{s_1} \times \underline{s_2} \times \cdots \times \underline{s_k}$ .

**5.90. Complexity of Binary Arithmetic.** Let  $x$  and  $y$  be  $k$ -bit numbers (this means the base-2 expansions of  $x$  and  $y$  have zeroes beyond the first  $k$  positions). (a) Show that there is an algorithm to compute the base-2 expansion of  $x + y$  (or  $x - y$ ) that requires at most  $ck$  bit operations, for some constant  $c$ . (b) Show that the base-2 expansion of  $xy$  can be computed in at most  $ck^2$  bit operations, for some constant  $c$ . (See 7.175 for a faster method.) (c) If  $y > 0$ , there exist unique  $q, r \in \mathbb{Z}$  with  $x = qy + r$  and  $0 \leq r < y$ . Show that  $q$  and  $r$  can be computed from  $x$  and  $y$  in at most  $ck^2$  bit operations, for some constant  $c$ .

**5.91. Binary Exponentiation.** Suppose  $n$  is a  $k$ -bit number,  $x \in \underline{n}$ , and  $e$  is an  $m$ -bit number. Show that we can compute  $x^e \bmod n$  in at most  $ck^2m$  bit operations, for some constant  $c$ .

**5.92. Euclid's Algorithm for Computing GCD's.** (a) Show that for all nonzero  $x \in \mathbb{Z}$ ,  $\gcd(x, 0) = x$ . Show that if  $x, y, q, r \in \mathbb{Z}$  satisfy  $y \neq 0$  and  $x = qy + r$  then  $\gcd(x, y) = \gcd(y, r)$ . (b) Use (a) and 5.85 to develop an algorithm that will compute the gcd of two  $n$ -bit numbers using at most  $n$  integer divisions (hence at most  $cn^3$  bit operations, for some constant  $c$ ).

**5.93.** (a) Devise a ranking algorithm for  $k$ -element multisets of an  $n$ -element ordered alphabet based on the recursion 2.26. Use this to rank the multiset  $[b, b, c, d, d, d]$  over the alphabet  $\{a, b, c, d, e\}$  and to unrank 132 to get a 6-element multiset over this alphabet. (b) Repeat part (a), but use a ranking algorithm based on one of the bijections in §1.11.

**5.94.** Fix  $k \in \mathbb{N}^+$ . Prove that every  $m \in \mathbb{N}$  can be written in exactly one way in the form  $m = \sum_{j=1}^k \binom{i_j}{j}$ , where  $0 \leq i_1 < i_2 < \cdots < i_k$ .



**5.95.** Fix  $k \in \mathbb{N}^+$ . Use 5.94 to find an explicit formula for a bijection  $f : \mathbb{N}^k \rightarrow \mathbb{N}$  (cf. 1.149).

**5.96.** (a) Devise a ranking algorithm for derangements based on the recursion 4.24. (b) List all derangements of  $\{1, 2, 3, 4\}$  in the order specified by your ranking algorithm. (c) Compute the rank of  $3527614 \in D_7$ . (d) Unrank 1776 to obtain a derangement in  $D_7$ .

**5.97.** Devise algorithms to rank and unrank partitions of  $n$  into  $k$  *distinct* parts. Use your algorithms to rank the partition  $(10, 7, 6, 3, 1)$  and unrank 10 to get a partition of 20 into 3 distinct parts.

**5.98.** Suppose we rewrite the recursion for Stirling numbers in the form

$$S(n, k) = S(n-1, k)k + S(n-1, k-1) \quad (n, k > 0).$$

(a) Use the bijective product and sum rules (taking terms in the order written here) to devise ranking and unranking algorithms for set partitions in  $SP(n, k)$ . (b) Rank the partition  $\pi = \{\{1, 7\}, \{2, 4, 5\}, \{3, 8\}, \{6\}\}$  and unrank 111 to obtain an element of  $SP(7, 3)$  (cf. 5.28). (c) Repeat 5.71 using the new ranking algorithm.

**5.99.** Use the recursion in 2.53 to develop ranking and unranking algorithms for the set  $SP(n)$  of all set partitions of an  $n$ -element set. Find the rank of  $\{\{1, 2, 4\}, \{3, 5, 6\}, \{7, 8\}\}$ . Which set partition of 8 has rank 1394?

**5.100.** Find ranking and unranking algorithms for three-of-a-kind poker hands. Use these algorithms to rank the three-of-a-kind hand  $\{4\heartsuit, 4\clubsuit, 4\diamondsuit, 9\clubsuit, K\spadesuit\}$  and to unrank 21,751.

**5.101.** Find ranking and unranking algorithms for one-pair poker hands. Use these algorithms to rank the one-pair hand  $\{2\heartsuit, 2\clubsuit, 7\diamondsuit, 8\diamondsuit, 10\diamondsuit\}$  and to unrank 497,079.

**5.102.** (a) Find ranking and unranking algorithms for straight poker hands (including straight flushes). Use these algorithms to rank the straight hand  $\{4\heartsuit, 5\heartsuit, 6\clubsuit, 7\diamondsuit, 8\diamondsuit\}$  and to unrank 1574. (b) Repeat part (a) for the set of straight hands that are not flushes.

**5.103.** (a) Find ranking and unranking algorithms for flush poker hands (including straight flushes). Use these algorithms to rank the flush hand  $\{3\heartsuit, 7\heartsuit, 10\heartsuit, J\heartsuit, K\heartsuit\}$  and to unrank 4716. (b) Repeat part (a) for the set of flush hands that are not straights.

**5.104.** Develop ranking and unranking algorithms for 231-avoiding permutations. Find the rank of  $w = 1\,5\,2\,4\,3\,11\,7\,6\,10\,8\,9$ . Unrank 231 to get a 231-avoiding permutation of length 7.

**5.105.** Develop ranking and unranking algorithms for the set of subsets of  $\{1, 2, \dots, n\}$  that do not contain two consecutive integers (cf. 2.130(b)). For  $n = 10$ , rank the subset  $\{2, 5, 7, 10\}$  and unrank 42.

**5.106.** (a) Use the pruning bijections in §3.12 to develop ranking and unranking algorithms for the set of trees with vertex set  $\{v_1, \dots, v_n\}$  such that  $\deg(v_i) = d_i$  for all  $i$  (where the  $d_i$  are positive integers summing to  $2n - 2$ ). (b) Given  $(d_1, \dots, d_9) = (1, 2, 1, 1, 1, 2, 3, 1, 4)$ , find the rank of the tree shown in Figure 3.16. (c) Unrank 129 to obtain a tree with the degrees  $d_i$  from part (b).

**5.107.** Write a successor algorithm for listing integer partitions of  $n$  into  $k$  parts in lexicographic order (see 10.36). Use your algorithm to find the successors of  $(9, 3)$ ,  $(7, 4, 2, 1)$ , and  $(3, 3, 1, 1, 1)$ .

**5.108.** Write a successor algorithm for listing all integer partitions of  $n$  in lexicographic order (see 10.36). Use your algorithm to find the successors of  $(9, 3)$ ,  $(7, 4, 2, 1)$ , and  $(3, 3, 1, 1, 1)$ .

**5.109.** Write a successor algorithm for listing set partitions of  $\{1, 2, \dots, n\}$  into  $k$  blocks, using any convenient ordering. Find the successor and predecessor of the set partition  $\{\{1, 7\}, \{2\}, \{3, 5, 6\}, \{4, 8\}\}$ .

**5.110.** (a) Write a successor algorithm for listing full-house poker hands, using any convenient ordering. (b) Use your algorithm to determine the successor of the hand  $\{J\clubsuit, J\diamondsuit, J\spadesuit, 9\clubsuit, 9\heartsuit\}$ . (c) What is the predecessor of the hand in (b)?

**5.111.** (a) Write a successor algorithm for listing one-pair poker hands, using any convenient ordering. (b) Use your algorithm to find the successor of the hand  $\{2\heartsuit, 2\clubsuit, 7\diamondsuit, 8\diamondsuit, 10\diamondsuit\}$ . (c) What is the predecessor of the hand in (b)?

**5.112.** Describe a successor algorithm for ranking rooted trees with vertex set  $\{1, 2, \dots, n\}$  rooted at vertex 1. Compute the successor and predecessor of the tree shown in Figure 3.9.

**5.113.** Devise a random selection algorithm for choosing anagrams in  $\mathcal{R}(a_1^{n_1} \cdots a_k^{n_k})$ .

**5.114.** (a) Devise a random selection algorithm for choosing integer partitions of  $n$  into  $k$  parts. (b) Write an algorithm for choosing a random integer partition of  $n$ .

**5.115.** Devise a random selection algorithm for choosing a derangement of  $n$  letters.

**5.116.** Confirm that the random selection algorithm for permutations described at the end of §5.14 will generate every permutation in  $S_n$  with equal probability.

**5.117.** Consider the following proposed algorithm for randomly selecting a permutation  $w \in S_n$ . Initially, set  $w_i = i$  for  $1 \leq i \leq n$ . Next, for  $1 \leq i \leq n$ , exchange  $w_i$  and  $w_j$ , where  $j$  is chosen randomly in  $\{1, 2, \dots, n\}$ . Does this method produce every element of  $S_n$  with equal probability? Explain.

**5.118.** Modify the algorithm in the previous exercise by exchanging  $w_i$  with  $w_j$ , where  $j$  is chosen randomly in  $\{1, 2, \dots, i\}$  at each stage. Does this method produce every element of  $S_n$  with equal probability? Explain.

**5.119.** Consider the following proposed algorithm for randomly selecting a permutation  $w \in S_n$ . Choose  $w_1 \in \{1, 2, \dots, n\}$  at random. For  $i = 2, \dots, n$  in turn, repeatedly choose  $w_i \in \{1, 2, \dots, n\}$  at random until a value different from  $w_1, \dots, w_{i-1}$  is obtained. Argue informally that this algorithm will produce every permutation in  $S_n$  with equal likelihood, but that the expected number of random choices needed to generate one permutation in  $S_n$  is

$$n(1 + 1/2 + 1/3 + \cdots + 1/n) \approx n \ln n.$$

**5.120.** Devise a ranking algorithm for 4-letter words in which Q is always followed by U (so Q cannot be the last letter). Use your algorithm to rank AQUA and QUIT and to unrank 1000. Can you find an algorithm that generates these words in alphabetical order? Can you generalize to  $n$ -letter words?

**5.121.** Devise a ranking algorithm for 5-letter words that never have two consecutive vowels. Use your algorithm to rank BILBO and THIRD and to unrank 9999. Can you find an algorithm that generates these words in alphabetical order? Can you generalize to  $n$ -letter words?

---

## Notes

Our presentation of ranking and random selection places great emphasis on bijections constructed automatically by repeated use of the bijective sum and product rules. For a somewhat different approach based on a multigraph model, see the papers [141, 142]. Other discussions of ranking and related problems can be found in the texts [10, 100, 128]. An encyclopedic treatment of algorithms for generating combinatorial objects may be found in Knuth's comprehensive treatise [78].

---

## Counting Weighted Objects

---

In earlier chapters, we have spent a lot of time studying the *counting problem*: given a finite set  $S$ , how many elements does  $S$  have? This chapter generalizes the counting problem to the following situation. Given a finite set  $S$  of objects, where each object is assigned an integer-valued *weight*, how many objects in  $S$  are there of each given weight? A convenient way to present the answer to this question is via a *generating function*, which is a polynomial  $a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$  such that  $a_k$  (the coefficient of  $x^k$ ) is the number of objects in  $S$  of weight  $k$ . After giving the basic definitions, we will develop rules for manipulating generating functions that are analogous to the sum rule and product rule from Chapter 1. We will also derive formulas for certain generating functions that generalize factorials, binomial coefficients, and multinomial coefficients. In later chapters, we extend all of these ideas to the more general situation where  $S$  is an *infinite* set of weighted objects.

---

### 6.1 Weighted Sets

This section presents the basic definitions needed to discuss sets of weighted objects, together with many examples.

**6.1. Definition: Weighted Sets.** A *weighted set* is a pair  $(S, \text{wt})$ , where  $S$  is a set and  $\text{wt} : S \rightarrow \mathbb{N}$  is a function from  $S$  to the nonnegative integers. For each  $z \in S$ , the integer  $\text{wt}(z)$  is called the *weight* of  $z$ . In this definition,  $S$  is not required to be finite, although we shall always make that assumption in this chapter. The weight function  $\text{wt}$  is also sometimes referred to as a *statistic* on  $S$ . If the weight function is understood from the context, we may sometimes refer to “the weighted set  $S$ .”

**6.2. Definition: Generating Function for a Weighted Set.** Given a finite weighted set  $(S, \text{wt})$ , the *generating function* for  $S$  is the polynomial

$$G_{S, \text{wt}}(x) = \sum_{z \in S} x^{\text{wt}(z)}.$$

We also write  $G_S(x)$  or  $G(x)$  if the weight function and set are understood from context. Note that the sum on the right side is well-defined, since  $S$  is finite and addition of polynomials is an associative and commutative operation (see the discussion following 2.2 and 2.149).

**6.3. Example.** Suppose  $S = \{a, b, c, d, e, f\}$ , and  $\text{wt} : S \rightarrow \mathbb{N}$  is given by

$$\text{wt}(a) = 4, \text{wt}(b) = 1, \text{wt}(c) = 0, \text{wt}(d) = 4, \text{wt}(e) = 4, \text{wt}(f) = 1.$$

The generating function for  $(S, \text{wt})$  is

$$G_{S, \text{wt}}(x) = x^{\text{wt}(a)} + x^{\text{wt}(b)} + \cdots + x^{\text{wt}(f)} = x^4 + x^1 + x^0 + x^4 + x^4 + x^1 = 1 + 2x + 3x^4.$$

Consider another weight function  $w : S \rightarrow \mathbb{N}$  given by  $w(a) = 0$ ,  $w(b) = 1$ ,  $w(c) = 2$ ,  $w(d) = 3$ ,  $w(e) = 4$ , and  $w(f) = 5$ . Using this weight function, we obtain a different generating function, namely

$$G_{S,w}(x) = 1 + x + x^2 + x^3 + x^4 + x^5.$$

**6.4. Example.** Suppose  $S$  is the set of all subsets of  $\{1, 2, 3\}$ , so

$$S = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Consider three different weight functions  $w_i : S \rightarrow \mathbb{N}$ , given by

$$w_1(A) = |A|; \quad w_2(A) = \sum_{i \in A} i; \quad w_3(A) = \min_{i \in A} i; \quad (\text{for all } A \in S).$$

(By convention, define  $w_3(\emptyset) = 0$ .) Each of these statistics leads to a different generating function:

$$\begin{aligned} G_{S,w_1}(x) &= x^0 + x^1 + x^1 + x^1 + x^2 + x^2 + x^2 + x^3 = 1 + 3x + 3x^2 + x^3 = (1 + x)^3; \\ G_{S,w_2}(x) &= x^0 + x^1 + x^2 + x^3 + x^3 + x^4 + x^5 + x^6 = 1 + x + x^2 + 2x^3 + x^4 + x^5 + x^6; \\ G_{S,w_3}(x) &= x^0 + x^1 + x^2 + x^3 + x^1 + x^1 + x^2 + x^1 = 1 + 4x + 2x^2 + x^3. \end{aligned}$$

**6.5. Example.** For each integer  $n \geq 0$ , consider the set  $\underline{n} = \{0, 1, 2, \dots, n-1\}$ . Define a weight function on this set by letting  $\text{wt}(i) = i$  for all  $i \in \underline{n}$ . The associated generating function is

$$G_{\underline{n}, \text{wt}}(x) = x^0 + x^1 + x^2 + \dots + x^{n-1} = \frac{x^n - 1}{x - 1}.$$

The last equality can be verified by using the distributive law to calculate

$$(x - 1)(1 + x + x^2 + \dots + x^{n-1}) = x^n - 1.$$

The generating function in this example will be a recurring building block in our later work, so we give it a special name.

**6.6. Definition: Quantum Integers.** If  $n$  is a positive integer and  $x$  is any variable, we define

$$[n]_x = 1 + x + x^2 + \dots + x^{n-1} = \frac{x^n - 1}{x - 1}.$$

We also define  $[0]_x = 0$ . The polynomial  $[n]_x$  is called the *quantum integer*  $n$  (relative to the variable  $x$ ).

**6.7. Example.** Let  $S$  be the set of all lattice paths from  $(0, 0)$  to  $(2, 3)$ . For  $P \in S$ , let  $w(P)$  be the number of unit squares in the region bounded by  $P$ , the  $x$ -axis, and the line  $x = 2$ . Let  $w'(P)$  be the number of unit squares in the region bounded by  $P$ , the  $y$ -axis, and the line  $y = 3$ . By examining the paths in Figure 1.7, we compute

$$\begin{aligned} G_{S,w}(x) &= x^6 + x^5 + x^4 + x^4 + x^3 + x^2 + x^3 + x^2 + x^1 + x^0 \\ &= 1 + x + 2x^2 + 2x^3 + 2x^4 + x^5 + x^6; \\ G_{S,w'}(x) &= x^0 + x^1 + x^2 + x^2 + x^3 + x^4 + x^3 + x^4 + x^5 + x^6 \\ &= 1 + x + 2x^2 + 2x^3 + 2x^4 + x^5 + x^6. \end{aligned}$$

Although the two weight functions are not equal (since there are paths  $P$  with  $w(P) \neq w'(P)$ ), it happens that  $G_{S,w} = G_{S,w'}$  in this example.

Now, consider the set  $T$  of Dyck paths from  $(0, 0)$  to  $(3, 3)$ . For  $P \in T$ , let  $\text{wt}(P)$  be the number of complete unit squares located between  $P$  and the diagonal line  $y = x$ . Using Figure 1.8, we find that

$$G_{T, \text{wt}}(x) = x^3 + x^2 + x^1 + x^1 + x^0 = 1 + 2x + x^2 + x^3.$$

**6.8. Remark.** Let  $(S, \text{wt})$  be a finite set of weighted objects. We know  $G_S(x) = \sum_{z \in S} x^{\text{wt}(z)}$ . By collecting together equal powers of  $x$  (as done in the calculations above), we can write  $G_S$  in the standard form

$$G_S(x) = a_0 x^0 + a_1 x^1 + a_2 x^2 + \cdots + a_m x^m \quad (a_i \in \mathbb{N}).$$

Comparing the two formulas for  $G_S$ , we see that *the coefficient  $a_i$  of  $x^i$  in  $G_{S, \text{wt}}(x)$  is the number of objects  $z$  in  $S$  such that  $\text{wt}(z) = i$* . We now illustrate this observation with several examples.

**6.9. Example.** Suppose  $T$  is the set of all set partitions of an  $n$ -element set, and the weight of a partition is the number of blocks in the partition. By definition of the Stirling number of the second kind (see 2.51), we have

$$G_T(x) = \sum_{k=0}^n S(n, k) x^k.$$

Similarly, if  $U$  is the set of all permutations of  $n$  elements, weighted by the number of cycles in the disjoint cycle decomposition, then

$$G_U(x) = \sum_{k=0}^n s'(n, k) x^k$$

where  $s'(n, k)$  is a signless Stirling number of the first kind (see §3.6). Finally, if  $V$  is the set of all integer partitions of  $n$ , weighted by number of parts, then

$$G_V(x) = \sum_{k=0}^n p(n, k) x^k.$$

This is also the generating function for  $V$  if we weight a partition by the length of its largest part (§2.8).

**6.10. Remark.** Suppose we replace the variable  $x$  in  $G_S(x)$  by the value 1. We obtain  $G_S(1) = \sum_{z \in S} 1^{\text{wt}(z)} = \sum_{z \in S} 1 = |S|$ . For example, in 6.7,  $G_{T, \text{wt}}(1) = 5 = C_3$ ; in 6.9,  $G_T(1) = B(n)$  (the Bell number),  $G_U(1) = n!$ , and  $G_V(1) = p(n)$ . Thus, the generating function  $G_S(x)$  can be viewed as a weighted analogue of “the number of elements in  $S$ .” On the other hand, using the convention that  $0^0 = 1$ ,  $G_S(0)$  is the number of objects in  $S$  having weight zero.

We also note that the polynomial  $G_S(x)$  can sometimes be factored or otherwise simplified, as illustrated by the first weight function in 6.4. Different statistics on  $S$  usually lead to different generating functions, but this is not always true (see 6.4 and 6.7).

Our goal in this chapter is to develop techniques for finding and manipulating generating functions that avoid listing all the objects in  $S$ , as we did in the examples above.

## 6.2 Inversions

Before presenting the sum and product rules for generating functions, we introduce an example of a weight function that arises frequently in algebraic combinatorics.

**6.11. Definition: Inversions.** Suppose  $w = w_1 w_2 \cdots w_n$  is a word, where each letter  $w_i$  is an integer. An *inversion* of  $w$  is a pair of indices  $i < j$  such that  $w_i > w_j$ . We write  $\text{inv}(w)$  for the number of inversions of  $w$ ; in other words,

$$\text{inv}(w_1 w_2 \cdots w_n) = \sum_{1 \leq i < j \leq n} \chi(w_i > w_j).$$

Thus  $\text{inv}(w)$  counts pairs of letters in  $w$  (not necessarily adjacent) that are out of numerical order. We also define  $\text{Inv}(w)$  to be the *set* of all inversion pairs  $(i, j)$ , so  $|\text{Inv}(w)| = \text{inv}(w)$ . If  $S$  is any finite set of words over the alphabet  $\mathbb{Z}$ , then

$$G_{S, \text{inv}}(x) = \sum_{w \in S} x^{\text{inv}(w)}$$

is the *inversion generating function* for  $S$ . These definitions extend to words over any *totally ordered* alphabet.

**6.12. Example.** Consider the word  $w = 414253$ ; here  $w_1 = 4$ ,  $w_2 = 1$ ,  $w_3 = 4$ , etc. The pair  $(1, 2)$  is an inversion of  $w$  since  $w_1 = 4 > 1 = w_2$ . The pair  $(2, 3)$  is not an inversion, since  $w_2 = 1 \leq 4 = w_3$ . Similarly,  $(1, 3)$  is not an inversion. Continuing in this way, we find that

$$\text{Inv}(w) = \{(1, 2), (1, 4), (1, 6), (3, 4), (3, 6), (5, 6)\},$$

so  $\text{inv}(w) = 6$ .

**6.13. Example.** Let  $S$  be the set of all permutations of  $\{1, 2, 3\}$ . We know that

$$S = \{123, 132, 213, 231, 312, 321\}.$$

Counting inversions, we conclude that

$$G_{S, \text{inv}}(x) = x^0 + x^1 + x^1 + x^2 + x^2 + x^3 = 1 + 2x + 2x^2 + x^3 = 1(1+x)(1+x+x^2).$$

Note that  $G_S(1) = 6 = 3! = |S|$ . Similarly, if  $T$  is the set of all permutations of  $\{1, 2, 3, 4\}$ , a longer calculation (see 6.50) leads to

$$G_{T, \text{inv}}(x) = 1 + 3x + 5x^2 + 6x^3 + 5x^4 + 3x^5 + x^6 = 1(1+x)(1+x+x^2)(1+x+x^2+x^3).$$

The factorization patterns in these examples will be explained and generalized below.

**6.14. Example.** Let  $S = \mathcal{R}(0^2 1^3)$  be the set of all rearrangements of two zeroes and three ones. We know that

$$S = \{00111, 01011, 01101, 01110, 10011, 10101, 10110, 11001, 11010, 11100\}.$$

Counting inversions, we conclude that

$$G_{S, \text{inv}}(x) = x^0 + x^1 + x^2 + x^3 + x^2 + x^3 + x^4 + x^4 + x^5 + x^6 = 1 + x + 2x^2 + 2x^3 + 2x^4 + x^5 + x^6.$$

The reader may notice that this is the same generating function that appeared in 6.7. This is not a coincidence; we explain why this happens in the next section.

**6.15. Example.** Let  $S = \mathcal{R}(a^1b^1c^2)$ , where we use  $a < b < c$  as the ordering of the alphabet. We know that

$$S = \{abcc, acbc, accb, bacc, bcac, bcca, cabc, cacb, cbac, cbca, ccab, ccba\}.$$

Counting inversions leads to

$$G_{S,\text{inv}}(x) = 1 + 2x + 3x^2 + 3x^3 + 2x^4 + x^5.$$

Now let  $T = \mathcal{R}(a^1b^2c^1)$  and  $U = \mathcal{R}(a^2b^1c^1)$  with the same ordering of the alphabet. The reader is invited to confirm that

$$G_{S,\text{inv}}(x) = G_{T,\text{inv}}(x) = G_{U,\text{inv}}(x),$$

although the sets of words in question are all different. This phenomenon will also be explained in the coming sections.

**6.16. Remark.** It can be shown that for any word  $w$ ,  $\text{inv}(w)$  is the minimum number of transpositions of adjacent letters required to sort the letters of  $w$  into weakly increasing order (see 9.29 and 9.179).

### 6.3 Weight-Preserving Bijections

In the next few sections, we introduce three fundamental rules that we will use to give combinatorial derivations of many generating function formulas. These rules are weighted analogues of the counting rules studied in Chapter 1. The first rule generalizes 1.30. We need one new definition to state this rule.

**6.17. Definition: Weight-Preserving Bijections.** Let  $(S, w_1)$  and  $(T, w_2)$  be two weighted sets. A *weight-preserving bijection* from  $(S, w_1)$  to  $(T, w_2)$  is a bijection  $f : S \rightarrow T$  such that

$$w_2(f(z)) = w_1(z) \quad \text{for all } z \in S.$$

**6.18. Theorem: Bijection Rule for Generating Functions.** Suppose  $(S, w_1)$  and  $(T, w_2)$  are two finite weighted sets such that there exists a weight-preserving bijection  $f : S \rightarrow T$ . Then

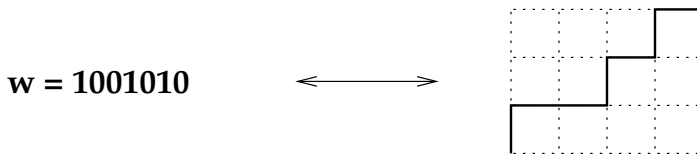
$$G_{S, w_1}(x) = G_{T, w_2}(x).$$

*Proof.* Let  $g : T \rightarrow S$  be the inverse of  $f$ . One verifies that  $g$  is a weight-preserving bijection, since  $f$  is. For each  $k \geq 0$ , let  $S_k = \{z \in S : w_1(z) = k\}$  and  $T_k = \{u \in T : w_2(u) = k\}$ . Since  $f$  and  $g$  preserve weights, they restrict to give maps  $f_k : S_k \rightarrow T_k$  and  $g_k : T_k \rightarrow S_k$  that are mutual inverses. Therefore,  $|S_k| = |T_k|$  for all  $k \geq 0$ . It follows that

$$G_{S, w_1}(x) = \sum_{k \geq 0} |S_k| x^k = \sum_{k \geq 0} |T_k| x^k = G_{T, w_2}(x). \quad \square$$

**6.19. Example.** Let  $S$  be the set of all lattice paths  $P$  from  $(0, 0)$  to  $(a, b)$ , and let  $\text{area}(P)$  be the area below the path and above the  $x$ -axis (cf. 6.7). Let  $T = \mathcal{R}(0^a 1^b)$  be the set of all words consisting of  $a$  zeroes and  $b$  ones, weighted by inversions. There is a bijection  $g : T \rightarrow S$  obtained by converting zeroes to east steps and ones to north steps. By examining a picture, one sees that  $\text{inv}(w) = \text{area}(g(w))$  for all  $w \in T$ . For example, if  $w = 1001010$ ,



**FIGURE 6.1**

Inversions of a word vs. area under a lattice path.

then  $g(w)$  is the lattice path shown in Figure 6.1. The four area cells in the lowest row correspond to the inversions between the first 1 in  $w$  and the four zeroes occurring later. Similarly, the two area cells in the next lowest row come from the inversions between the second 1 in  $w$  and the two zeroes occurring later. Since  $g$  is a weight-preserving bijection, we conclude that

$$G_{T,\text{inv}}(x) = G_{S,\text{area}}(x).$$

When  $a = 2$  and  $b = 3$ , this explains the equality of generating functions observed in 6.7 and 6.14. In 6.7, we also considered another weight on paths  $P \in S$ , namely the number of area squares between  $P$  and the  $y$ -axis. Denoting this weight by  $\text{area}'$ , we have

$$G_{S,\text{area}}(x) = G_{S,\text{area}'}(x)$$

(for arbitrary  $a$  and  $b$ ). This follows from the weight-preserving bijection rule, since rotating a path  $180^\circ$  about  $(a/2, b/2)$  defines a bijection  $r : S \rightarrow S$  that sends  $\text{area}$  to  $\text{area}'$ . Similarly, letting  $S'$  be the set of paths from  $(0, 0)$  to  $(b, a)$ , we have

$$G_{S,\text{area}}(x) = G_{S',\text{area}'}(x) \quad (= G_{S',\text{area}}(x))$$

since reflection in the diagonal line  $y = x$  defines a weight-preserving bijection from  $(S, \text{area})$  to  $(S', \text{area}')$ . We will soon derive an explicit formula for the generating functions occurring in this example (§6.7).

## 6.4 Sum and Product Rules for Weighted Sets

Now we discuss the weighted analogues of the sum and product rules.

**6.20. Theorem: Sum Rule for Weighted Sets.** Suppose  $(S_1, \text{wt}_1)$ ,  $(S_2, \text{wt}_2)$ ,  $\dots$ ,  $(S_k, \text{wt}_k)$  are finite weighted sets such that  $S_1, \dots, S_k$  are pairwise disjoint sets. Let  $S = S_1 \cup \dots \cup S_k$ , and define  $\text{wt} : S \rightarrow \mathbb{N}$  by setting  $\text{wt}(z) = \text{wt}_i(z)$  for all  $z \in S_i$ . Then

$$G_{S,\text{wt}}(x) = G_{S_1,\text{wt}_1}(x) + G_{S_2,\text{wt}_2}(x) + \dots + G_{S_k,\text{wt}_k}(x).$$

*Proof.* By definition,  $G_{S,\text{wt}}(x) = \sum_{z \in S} x^{\text{wt}(z)}$ . Because addition of polynomials is commutative and associative, we can order the terms of this sum so that all objects in  $S_1$  come first, followed by objects in  $S_2$ , and so on, ending with all objects in  $S_k$ . We obtain

$$\begin{aligned} G_{S,\text{wt}}(x) &= \sum_{z \in S_1} x^{\text{wt}(z)} + \sum_{z \in S_2} x^{\text{wt}(z)} + \dots + \sum_{z \in S_k} x^{\text{wt}(z)} \\ &= \sum_{z \in S_1} x^{\text{wt}_1(z)} + \sum_{z \in S_2} x^{\text{wt}_2(z)} + \dots + \sum_{z \in S_k} x^{\text{wt}_k(z)} \\ &= G_{S_1,\text{wt}_1}(x) + G_{S_2,\text{wt}_2}(x) + \dots + G_{S_k,\text{wt}_k}(x). \quad \square \end{aligned}$$

**6.21. Theorem: Product Rule for Two Weighted Sets.** Suppose  $(T, w_1)$  and  $(U, w_2)$  are two finite weighted sets. On the product set  $S = T \times U$ , define a weight  $w$  by setting  $w((t, u)) = w_1(t) + w_2(u)$  for  $t \in T$  and  $u \in U$ . Then

$$G_{S,w}(x) = G_{T,w_1}(x) \cdot G_{U,w_2}(x).$$

*Proof.* The proof consists of the following calculation, which requires the generalized distributive law (in the form given by 2.5):

$$\begin{aligned} G_{S,w}(x) &= \sum_{(t,u) \in S} x^{w((t,u))} = \sum_{(t,u) \in T \times U} x^{w_1(t) + w_2(u)} \\ &= \sum_{(t,u) \in T \times U} x^{w_1(t)} \cdot x^{w_2(u)} \\ &= \left( \sum_{t \in T} x^{w_1(t)} \right) \cdot \left( \sum_{u \in U} x^{w_2(u)} \right) \\ &= G_{T,w_1}(x) \cdot G_{U,w_2}(x). \quad \square \end{aligned}$$

**6.22. Theorem: Product Rule for  $k$  Weighted Sets.** Suppose that  $(S_i, w_i)$  is a finite weighted set for  $1 \leq i \leq k$ . On the product set  $S = S_1 \times S_2 \times \cdots \times S_k$ , define a weight  $w$  by  $w((z_1, \dots, z_k)) = \sum_{i=1}^k w_i(z_i)$  for all  $z_i \in S_i$ . Then

$$G_{S,w}(x) = \prod_{i=1}^k G_{S_i, w_i}(x).$$

*Proof.* This formula follows from the product rule for two weighted sets by induction on  $k$ . Alternatively, one can mimic the proof given above for two sets, this time using the full-blown version of the generalized distributive law (see 2.6).  $\square$

When discussing a Cartesian product of weighted sets, we always use the weight function on the product set given in the statement of the product rule (i.e., the weight is the sum of the weights of the component objects) unless otherwise specified.

**6.23. Example.** Given a positive integer  $n$ , consider the  $n$  weighted sets  $S_i = \underline{i} = \{0, 1, \dots, i-1\}$  (for  $1 \leq i \leq n$ ), where  $\text{wt}(z) = z$  for each  $z \in \underline{i}$ . We have seen in 6.5 that  $G_{S_i}(x) = [i]_x$ , the quantum integer  $i$ . Now consider the product set  $S = S_1 \times S_2 \times \cdots \times S_n = \underline{1} \times \underline{2} \times \cdots \times \underline{n}$ . By the product rule, the generating function for  $S$  is

$$G_S(x) = \prod_{i=1}^n [i]_x.$$

Since  $G_S(1) = |S| = n!$ , the polynomial  $G_S(x)$  is a weighted analogue of a factorial. This polynomial will arise frequently, so we give it a special name.

**6.24. Definition: Quantum Factorials.** For each  $n \geq 1$  and every variable  $x$ , define the *quantum factorial of  $n$  relative to  $x$*  to be the polynomial

$$[n]!_x = \prod_{i=1}^n [i]_x = \prod_{i=1}^n (1 + x + x^2 + \cdots + x^{i-1}) = \prod_{i=1}^n \frac{x^i - 1}{x - 1}.$$

Also define  $[0]!_x = 1$ .

Observe that  $[n]!_x = [n-1]!_x [n]_x$  for all  $n \geq 1$ .

**6.25. Example.** We have  $[0]!_x = 1 = [1]!_x$ ;  $[2]!_x = [2]_x = 1 + x$ ;

$$[3]!_x = (1 + x)(1 + x + x^2) = 1 + 2x + 2x^2 + x^3;$$

$$[4]!_x = (1 + x)(1 + x + x^2)(1 + x + x^2 + x^3) = 1 + 3x + 5x^2 + 6x^3 + 5x^4 + 3x^5 + x^6;$$

$$[5]!_x = 1 + 4x + 9x^2 + 15x^3 + 20x^4 + 22x^5 + 20x^6 + 15x^7 + 9x^8 + 4x^9 + x^{10}.$$

We can use other variables besides  $x$ ; for instance,  $[3]!_q = 1 + 2q + 2q^2 + q^3$ . Sometimes we will replace the variable here by a specific integer or real number; then the quantum factorial will evaluate to some specific number. For example, when  $q = 4$ ,  $[3]!_q = 1 + 8 + 32 + 64 = 105$ . As another example, when  $x = 1$ ,  $[n]!_x = n!$ .

## 6.5 Inversions and Quantum Factorials

The reader may have recognized some of the quantum factorial polynomials above as matching the inversion generating functions for permutations in 6.13. The next theorem proves that this pattern holds in general.

**6.26. Theorem: Quantum Factorials and Inversions.** For every  $n \geq 0$ , let  $S_n$  be the set of all permutations of  $\{1, 2, \dots, n\}$ , weighted by inversions. Then

$$G_{S_n, \text{inv}}(x) = \sum_{w \in S_n} x^{\text{inv}(w)} = [n]!_x.$$

*Proof.* Let  $T_n = \underline{1} \times \underline{2} \times \dots \times \underline{n}$  with the usual product weight; we saw in 6.23 that  $G_{T_n, \text{wt}}(x) = [n]!_x$ . Therefore, to prove the theorem, it suffices to define a weight-preserving bijection  $f_n : S_n \rightarrow T_n$ .

Let  $w = w_1 w_2 \dots w_n \in S_n$  be a permutation of  $\{1, 2, \dots, n\}$ . For each  $k$  between 1 and  $n$ , define  $t_k$  to be the number of pairs  $i < j$  such that  $w_i > w_j$  and  $w_i = k$ ; then define  $f_n(w) = (t_1, t_2, \dots, t_n)$ . In other words,  $t_k = |\{(i, j) \in \text{Inv}(w) : w_i = k\}|$  is the number of inversions that the symbol  $k$  has with smaller symbols to its right. There are  $k - 1$  possible symbols less than  $k$ , so  $t_k \in \{0, 1, 2, \dots, k - 1\} = \underline{k}$  for every  $k$ . Thus,  $f_n(w)$  does lie in the set  $T_n$ . For example, if  $w = 4, 2, 8, 5, 1, 6, 7, 3$ , then  $f_n(w) = (0, 1, 0, 3, 2, 1, 1, 5)$ . We have  $t_5 = 2$ , for instance, because of the two inversions  $(4, 5)$  and  $(4, 8)$  caused by the entries  $5 > 1$  and  $5 > 3$  in  $w$ . Every inversion of  $w$  is counted by exactly one of the numbers  $t_k$ , so that  $\text{inv}(w) = \sum_{k=1}^n t_k = \text{wt}(f_n(w))$  for all  $w \in S_n$ . This shows that  $f_n$  is a weight-preserving map.

To show that  $f_n$  is a bijection, we display a two-sided inverse map  $g_n : T_n \rightarrow S_n$ . We define  $g_n$  by means of a recursive *insertion procedure*. The cases  $n \leq 1$  are immediate since the sets involved have only one element. Assume  $n > 1$  and  $g_{n-1}$  has already been defined. Given  $(t_1, \dots, t_n) \in T_n = T_{n-1} \times \underline{n}$ , begin by computing  $v = g_{n-1}(t_1, \dots, t_{n-1})$ , which is a permutation of  $\{1, 2, \dots, n - 1\}$ . To find  $g_n(t_1, \dots, t_n)$ , we insert the symbol  $n$  into the permutation  $v$  in such a way that  $n$  will cause  $t_n$  new inversions. This can always be done in a unique way. For, there are  $n$  possible positions in  $v$  where the symbol  $n$  could be inserted (after the last letter, or immediately before one of the  $n - 1$  existing letters). If we insert  $n$  after the last letter, it will create no new inversions. Scanning to the left, if we insert  $n$  immediately before the  $k$ th letter from the far right, then this insertion will cause exactly  $k$  new inversions, since  $n$  exceeds all letters to its right, and no letter to the left of  $n$  exceeds  $n$ . Thus, the different insertion positions for  $n$  lead to  $0, 1, \dots$ , or  $n - 1$  new inversions,

and this is exactly the range of values for  $t_n$ . The recursive construction ensures that, for all  $k \leq n$ ,  $t_k$  is the number of inversion pairs involving the symbol  $k$  on the left and some smaller symbol on the right. One may check that  $g_n$  is the two-sided inverse of  $f_n$ .

Here is an iterative description of the computation of  $g_n(t_1, \dots, t_n)$ . Beginning with an empty word, successively insert  $1, 2, \dots, n$ . At stage  $k$ , insert  $k$  in the unique position that will increase the total inversion count by  $t_k$ . For example, let us compute  $g_8(0, 0, 1, 3, 2, 1, 5, 5)$ . In the first two steps, we generate  $1, 2$ , which has zero inversions. Then we place the  $3$  one position left of the far right slot, obtaining the permutation  $1, 3, 2$  with one inversion. Next we count three positions from the far right (arriving at the far left), and insert  $4$  to obtain the permutation  $4, 1, 3, 2$  with three new inversions and four total inversions. The process continues, leading to  $4, 1, 5, 3, 2$ ; then  $4, 1, 5, 3, 6, 2$ ; then  $4, 7, 1, 5, 3, 6, 2$ ; and finally to  $w = 4, 7, 8, 1, 5, 3, 6, 2$ . The reader may check that  $f_8(w) = (0, 0, 1, 3, 2, 1, 5, 5)$ .  $\square$

Because of the previous proof, we sometimes call the elements of  $T_n$  *inversion tables*. Other types of inversion tables for permutations can be constructed by classifying the inversions of  $w$  in different ways. For example, our proof classified inversions by the value of the leftmost symbol in the inversion pair. One can also classify inversions using the value of the rightmost symbol, the position of the leftmost symbol, or the position of the rightmost symbol. These possibilities are explored in the exercises.

## 6.6 Descents and Major Index

This section introduces more statistics on words, which will lead to another combinatorial interpretation for the quantum factorial  $[n]!_x$ .

**6.27. Definition: Descents and Major Index.** Let  $w = w_1 w_2 \cdots w_n$  be a word over a totally ordered alphabet  $A$ . The *descent set* of  $w$ , denoted  $\text{Des}(w)$ , is the set of all  $i < n$  such that  $w_i > w_{i+1}$ . This is the set of *positions* in  $w$  where a letter is immediately followed by a smaller letter. Define the *descent count* for  $w$  by  $\text{des}(w) = |\text{Des}(w)|$ . Define the *major index* of  $w$ , denoted  $\text{maj}(w)$ , to be the sum of the elements of the set  $\text{Des}(w)$ . In symbols, we can write

$$\text{des}(w) = \sum_{i=1}^{n-1} \chi(w_i > w_{i+1}); \quad \text{maj}(w) = \sum_{i=1}^{n-1} i \chi(w_i > w_{i+1}).$$

**6.28. Example.** If  $w = 47815362$ , then  $\text{Des}(w) = \{3, 5, 7\}$ ,  $\text{des}(w) = 3$ , and  $\text{maj}(w) = 3 + 5 + 7 = 15$ . If  $w = 101100101$ , then  $\text{Des}(w) = \{1, 4, 7\}$ ,  $\text{des}(w) = 3$ , and  $\text{maj}(w) = 12$ . If  $w = 33555789$ , then  $\text{Des}(w) = \emptyset$ ,  $\text{des}(w) = 0$ , and  $\text{maj}(w) = 0$ .

**6.29. Theorem: Quantum Factorials and Major Index.** For every  $n \geq 0$ , let  $S_n$  be the set of all permutations of  $\{1, 2, \dots, n\}$ , weighted by major index. Then  $G_{S_n, \text{maj}}(x) = \sum_{w \in S_n} x^{\text{maj}(w)} = [n]!_x$ .

*Proof.* As in the case of inversions, it suffices to define a weight-preserving bijection  $f_n : S_n \rightarrow T_n = \underline{1} \times \underline{2} \times \cdots \times \underline{n}$ . We use a variation of the “inversion-table” idea, adapted to the major index statistic. Given  $w \in S_n$  and  $0 \leq k \leq n$ , let  $w^{(k)}$  be the word obtained from  $w$  by erasing all letters larger than  $k$ . Then define  $f_n(w) = (t_1, t_2, \dots, t_n)$ , where  $t_k = \text{maj}(w^{(k)}) - \text{maj}(w^{(k-1)})$ . Intuitively, if we imagine building up  $w$  from the empty word by inserting  $1, 2, \dots, n$  in this order, then  $t_k$  records the extra contribution to  $\text{maj}$

caused by the insertion of the new symbol  $k$ . For example, given  $w = 42851673$ , we compute  $\text{maj}(\epsilon) = 0$ ,  $\text{maj}(1) = 0$ ,  $\text{maj}(21) = 1$ ,  $\text{maj}(213) = 1$ ,  $\text{maj}(4213) = 3$ ,  $\text{maj}(42513) = 4$ ,  $\text{maj}(425163) = 9$ ,  $\text{maj}(4251673) = 10$  and finally  $\text{maj}(w) = 15$ . It follows that  $f_8(w) = (0, 1, 0, 2, 1, 5, 1, 5)$ . Observe that  $\sum_{k=1}^n t_k = \text{maj}(w^{(n)}) - \text{maj}(w^{(0)}) = \text{maj}(w)$ , so the map  $f_n$  is weight-preserving.

We see from the definition that  $f_n(w) = (f_{n-1}(w^{(n-1)}), t_n = \text{maj}(w) - \text{maj}(w^{(n-1)}))$ . To show that  $f_n(w)$  does lie in  $T_n$ , it suffices by induction to show that  $t_n \in \underline{n} = \{0, 1, 2, \dots, n-1\}$ . Let us first consider an example. Suppose we wish to insert the new symbol 8 into the permutation  $w' = 4 > 2, 5 > 1, 6, 7 > 3$ , which satisfies  $\text{maj}(w') = 1 + 3 + 6 = 10$ . There are eight gaps into which the symbol 8 might be placed. Let us compute the major index of each of the resulting permutations:

$$\begin{array}{llll} \text{maj}(8 > 4 > 2, 5 > 1, 6, 7 > 3) & = 1 + 2 + 4 + 7 & = 14 = & \text{maj}(w') + 4; \\ \text{maj}(4, 8 > 2, 5 > 1, 6, 7 > 3) & = 2 + 4 + 7 & = 13 = & \text{maj}(w') + 3; \\ \text{maj}(4 > 2, 8 > 5 > 1, 6, 7 > 3) & = 1 + 3 + 4 + 7 & = 15 = & \text{maj}(w') + 5; \\ \text{maj}(4 > 2, 5, 8 > 1, 6, 7 > 3) & = 1 + 4 + 7 & = 12 = & \text{maj}(w') + 2; \\ \text{maj}(4 > 2, 5 > 1, 8 > 6, 7 > 3) & = 1 + 3 + 5 + 7 & = 16 = & \text{maj}(w') + 6; \\ \text{maj}(4 > 2, 5 > 1, 6, 8 > 7 > 3) & = 1 + 3 + 6 + 7 & = 17 = & \text{maj}(w') + 7; \\ \text{maj}(4 > 2, 5 > 1, 6, 7, 8 > 3) & = 1 + 3 + 7 & = 11 = & \text{maj}(w') + 1; \\ \text{maj}(4 > 2, 5 > 1, 6, 7 > 3, 8) & = 1 + 3 + 6 & = 10 = & \text{maj}(w') + 0. \end{array}$$

Observe that the possible values of  $t_n$  are precisely  $0, 1, 2, \dots, 7$  (in some order).

To see that this always happens, suppose  $w^{(n-1)}$  has descents at positions  $i_1 > i_2 > \dots > i_d$ , where  $1 \leq i_j \leq n-2$  for all  $i_j$ . There are  $n$  gaps between the  $n-1$  letters in  $w^{(n-1)}$ , including the positions at the far left and far right ends. Let us number these gaps  $0, 1, 2, \dots, n-1$  as follows. The gap at the far right end is numbered zero. Next, the gaps immediately to the right of the descents are numbered  $1, 2, \dots, d$  from right to left. (We have chosen the indexing of the descent positions so that the gap between positions  $i_j$  and  $i_j + 1$  receives the number  $j$ .) Then, the remaining gaps are numbered  $d+1, \dots, n-1$  starting at the far left end and working to the right. In the example considered above, the gaps would be numbered as follows:

$$\begin{array}{cccccccc} . & 4 & > & 2 & , & 5 & > & 1 & , & 6 & , & 7 & > & 3 & . \\ 4 & & 3 & & 5 & & 2 & & 6 & & 7 & & 1 & & 0 \end{array}$$

Note that inserting the symbol 8 into the gap labeled  $j$  causes the major index to increase by exactly  $j$ .

Let us prove that this happens in the general case. If we insert  $n$  into the far right gap of  $w^{(n-1)}$  (which is labeled zero), there will be no new descents, so  $\text{maj}(w^{(n)}) = \text{maj}(w^{(n-1)}) + 0$  as desired. Suppose we insert  $n$  into the gap labeled  $j$ , where  $1 \leq j \leq d$ . In  $w^{(n-1)}$ , we had  $w_{i_j} > w_{i_j+1}$ , but the insertion of  $n$  changes this configuration to  $w_{i_j} < n > w_{i_j+1}$ . This pushes the descent at position  $i_j$  one position to the right. Furthermore, the descents that formerly occurred at positions  $i_{j-1}, \dots, i_1$  (which are to the right of  $i_j$ ) also get pushed one position to the right because of the new symbol  $n$ . It follows that the major index increases by exactly  $j$ , as desired. Finally, suppose  $d < j \leq n-1$ . Let the gap labeled  $j$  occur at position  $u$  in the new word, and let  $t$  be the number of descents in the old word preceding this gap. By definition of the gap labeling, we must have  $j = (u - t) + d$ . On the other hand, inserting  $n$  in this gap produces a new descent at position  $u$ , and pushes the  $(d - t)$  descents located to the right of position  $u$  one position further right. The net change to the major index is therefore  $u + (d - t) = j$ , as desired.

We now know that  $t_k \in \{0, 1, 2, \dots, k-1\}$  for all  $k$ , so that  $f_n$  does map  $S_n$  into  $T_n$ . The preceding discussion also tells us how to invert the action of  $f_n$ . Given  $(t_1, \dots, t_n) \in T_n$ ,

we use the  $t_k$ 's to insert the numbers  $1, \dots, n$  into an initially empty permutation. After  $1, \dots, k-1$  have been inserted, we number the gaps according to the rules above and then insert  $k$  in the unique gap labeled  $t_k$ . We just proved that this insertion will increase the major index by  $t_k$ . It follows that the resulting permutation  $w \in S_n$  is the unique object satisfying  $f_n(w) = (t_1, \dots, t_n)$ . Therefore,  $f_n$  is a bijection.  $\square$

**6.30. Example.** Let us compute  $f_6^{-1}(0, 1, 1, 0, 4, 3)$ . Using the insertion algorithm from the preceding proof, we generate the following sequence of permutations: 1; then 2, 1; then 2, 3, 1; then 2, 3, 1, 4; then 2, 3, 1, 5, 4; and finally 6, 2, 3, 1, 5, 4.

## 6.7 Quantum Binomial Coefficients

The formula  $[n]!_x = \prod_{i=1}^n [i]_x$  for the quantum factorial is analogous to the formula  $n! = \prod_{i=1}^n i$  for the ordinary factorial. We can extend this analogy to binomial coefficients and multinomial coefficients. This leads to the following definitions.

**6.31. Definition: Quantum Binomial Coefficients.** Suppose  $n \geq 0$ ,  $0 \leq k \leq n$ , and  $x$  is any variable. We define the *quantum binomial coefficients* by the formula

$$\begin{bmatrix} n \\ k \end{bmatrix}_x = \frac{[n]!_x}{[k]!_x [n-k]!_x} = \frac{(x^n - 1)(x^{n-1} - 1) \cdots (x - 1)}{(x^k - 1)(x^{k-1} - 1) \cdots (x - 1)(x^{n-k} - 1)(x^{n-k-1} - 1) \cdots (x - 1)}.$$

**6.32. Definition: Quantum Multinomial Coefficients.** Suppose  $n_1, \dots, n_k \geq 0$  and  $x$  is any variable. We define the *quantum multinomial coefficients* by the formula

$$\begin{bmatrix} n_1 + \cdots + n_k \\ n_1, \dots, n_k \end{bmatrix}_x = \frac{[n_1 + \cdots + n_k]!_x}{[n_1]!_x [n_2]!_x \cdots [n_k]!_x} = \frac{(x^{n_1 + \cdots + n_k} - 1) \cdots (x^2 - 1)(x - 1)}{\prod_{i=1}^k [(x^{n_i} - 1)(x^{n_i-1} - 1) \cdots (x - 1)]}.$$

One cannot immediately see from the defining formulas that the quantum binomial and multinomial coefficients are actually polynomials in  $x$  (as opposed to quotients of polynomials). However, we will soon prove that these entities are polynomials with nonnegative integer coefficients. We will also give several combinatorial interpretations for quantum binomial and multinomial coefficients in terms of suitable weighted sets of objects. Before doing so, we need to develop a few more tools.

It is immediate from the definitions that  $\begin{bmatrix} n \\ k \end{bmatrix}_x = \begin{bmatrix} n \\ n-k \end{bmatrix}_x = \begin{bmatrix} n \\ k, n-k \end{bmatrix}_x$ ; in particular, quantum binomial coefficients are special cases of quantum multinomial coefficients. We will usually prefer to use multinomial coefficients, writing  $\begin{bmatrix} a+b \\ a, b \end{bmatrix}_x$  in preference to  $\begin{bmatrix} a+b \\ a \end{bmatrix}_x$  or  $\begin{bmatrix} a+b \\ b \end{bmatrix}_x$ , because in most combinatorial applications the parameters  $a$  and  $b$  are more natural than  $a$  and  $a+b$ .

Before entering into further combinatorial discussions, we pause to give an algebraic proof of two fundamental recursions satisfied by the quantum binomial coefficients. These recursions are both “quantum analogues” of the binomial coefficient recursion 2.25.

**6.33. Theorem: Recursions for Quantum Binomial Coefficients.** For all  $a, b > 0$ , we have the recursion

$$\begin{aligned} \begin{bmatrix} a+b \\ a, b \end{bmatrix}_x &= x^b \begin{bmatrix} a+b-1 \\ a-1, b \end{bmatrix}_x + \begin{bmatrix} a+b-1 \\ a, b-1 \end{bmatrix}_x \\ &= \begin{bmatrix} a+b-1 \\ a-1, b \end{bmatrix}_x + x^a \begin{bmatrix} a+b-1 \\ a, b-1 \end{bmatrix}_x. \end{aligned}$$

The initial conditions are  $\begin{bmatrix} a \\ a, 0 \end{bmatrix}_x = \begin{bmatrix} b \\ 0, b \end{bmatrix}_x = 1$ .

*Proof.* We prove the first equality, leaving the second one as an exercise. Writing out the definitions, the right side of the first recursion is

$$x^b \begin{bmatrix} a+b-1 \\ a-1, b \end{bmatrix}_x + \begin{bmatrix} a+b-1 \\ a, b-1 \end{bmatrix}_x = \frac{x^b [a+b-1]!_x}{[a-1]!_x [b]!_x} + \frac{[a+b-1]!_x}{[a]!_x [b-1]!_x}.$$

Multiply the first fraction by  $[a]_x/[a]_x$  and the second fraction by  $[b]_x/[b]_x$  to create a common denominator. Bringing out common factors, we obtain

$$\left( \frac{[a+b-1]!_x}{[a]!_x [b]!_x} \right) \cdot (x^b [a]_x + [b]_x).$$

By definition of quantum integers,

$$\begin{aligned} [b]_x + x^b [a]_x &= (1 + x + x^2 + \cdots + x^{b-1}) + x^b (1 + x + \cdots + x^{a-1}) \\ &= (1 + \cdots + x^{b-1}) + (x^b + \cdots + x^{a+b-1}) = [a+b]_x. \end{aligned}$$

Putting this into the previous formula, we get

$$\frac{[a+b-1]!_x [a+b]_x}{[a]!_x [b]!_x} = \begin{bmatrix} a+b \\ a, b \end{bmatrix}_x.$$

The initial conditions follow immediately from the definitions.  $\square$

**6.34. Corollary: Polynomiality of Quantum Binomial Coefficients.** For all  $n \geq 0$  and  $0 \leq k \leq n$ ,  $\begin{bmatrix} n \\ k \end{bmatrix}_x$  is a polynomial in  $x$  with nonnegative integer coefficients.

*Proof.* Use induction on  $n \geq 0$ , the base case  $n = 0$  being evident. Assume  $n > 0$  and that  $\begin{bmatrix} n-1 \\ j \end{bmatrix}_x$  is already known to be a polynomial with coefficients in  $\mathbb{N}$  for  $0 \leq j \leq n-1$ . Then, by the recursion derived above (taking  $n = a+b$  and  $k = a$ , so  $b = n-k$ ),

$$\begin{bmatrix} n \\ k \end{bmatrix}_x = x^{n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_x + \begin{bmatrix} n-1 \\ k \end{bmatrix}_x.$$

Thanks to the induction hypothesis, we know that the right side is a polynomial with coefficients in  $\mathbb{N}$ . This completes the induction argument.  $\square$

We need one more result before describing the combinatorial interpretations for the quantum binomial coefficients. This result can be viewed as a generalization of the weighted bijection rule.

**6.35. Theorem: Weight-Shifting Rule.** Suppose  $(S, w_1)$  and  $(T, w_2)$  are finite weighted sets, and  $f : S \rightarrow T$  is a bijection such that, for some constant  $b$ ,  $w_1(z) = w_2(f(z)) + b$  holds for all  $z \in S$ . Then

$$G_{S, w_1}(x) = x^b G_{T, w_2}(x).$$

*Proof.* Let  $\{\star\}$  be a one-point set with  $\text{wt}(\star) = b$ . The generating function for this weighted set is  $x^b$ . There is a bijection  $i : T \rightarrow T \times \{\star\}$  given by  $t \mapsto (t, \star)$  for  $t \in T$ . Observe that  $i \circ f : S \rightarrow T \times \{\star\}$  is a *weight-preserving* bijection, since

$$w_1(z) = w_2(f(z)) + b = w_2(f(z)) + \text{wt}(\star) = \text{wt}((f(z), \star)) = \text{wt}(i(f(z))) \quad (z \in S).$$

Therefore, by the bijection rule and product rule for weighted sets,

$$G_S(x) = G_{T \times \{\star\}}(x) = G_T(x) G_{\{\star\}}(x) = x^b G_T(x). \quad \square$$

**6.36. Theorem: Combinatorial Interpretations of Quantum Binomial Coefficients.** Fix integers  $a, b \geq 0$ . Let  $L(a, b)$  be the set of all lattice paths from  $(0, 0)$  to  $(a, b)$ . Let  $P(a, b)$  be the set of integer partitions  $\mu$  with largest part at most  $a$  and with at most  $b$  parts. Then

$$\left[ \begin{matrix} a+b \\ a, b \end{matrix} \right]_x = \sum_{w \in \mathcal{R}(0^a 1^b)} x^{\text{inv}(w)} = \sum_{\pi \in L(a, b)} x^{\text{area}(\pi)} = \sum_{\pi \in L(a, b)} x^{\text{area}'(\pi)} = \sum_{\mu \in P(a, b)} x^{|\mu|}.$$

*Proof.* In 6.19, we constructed weight-preserving bijections between the three weighted sets  $(\mathcal{R}(0^a 1^b), \text{inv})$ ,  $(L(a, b), \text{area})$ , and  $(L(a, b), \text{area}')$ . Furthermore, Figure 2.18 shows that the weighted set  $(P(a, b), |\cdot|)$  is essentially identical to the weighted set  $(L(a, b), \text{area}')$ . So all of the combinatorial summations in the theorem are equal by the weighted bijection rule. We must prove that these all equal  $\left[ \begin{matrix} a+b \\ a, b \end{matrix} \right]_x$ . We give two proofs illustrating different techniques.

*First Proof.* For each  $a, b \geq 0$ , let  $g(a, b) = \sum_{\pi \in L(a, b)} x^{\text{area}(\pi)} = G_{L(a, b), \text{area}}(x)$ . Suppose we can show that this function satisfies the same recursion and initial conditions as the quantum binomial coefficients do, namely

$$g(a, b) = x^b g(a-1, b) + g(a, b-1); \quad g(a, 0) = g(0, b) = 1.$$

Then a routine induction on  $a+b$  will prove that  $\left[ \begin{matrix} a+b \\ a, b \end{matrix} \right]_x = g(a, b)$  for all  $a, b \geq 0$ .

To check the initial conditions, note that there is only one lattice path from  $(0, 0)$  to  $(a, 0)$ , and the area underneath this path is zero. So  $g(a, 0) = x^0 = 1$ . Similarly,  $g(0, b) = 1$ . Now let us prove the recursion for  $g(a, b)$ , assuming  $a, b > 0$ . The set  $L(a, b)$  is the disjoint union of sets  $L_1$  and  $L_2$ , where  $L_1$  consists of all paths from  $(0, 0)$  to  $(a, b)$  ending in an east step, and  $L_2$  consists of all paths from  $(0, 0)$  to  $(a, b)$  ending in a north step. See Figure 6.2. Deleting the final north step from a path in  $L_2$  defines a bijection from  $L_2$  to  $L(a, b-1)$ , which is weight-preserving since the area below the path is not affected by the deletion of the north step. It follows that  $G_{L_2, \text{area}}(x) = G_{L(a, b-1), \text{area}}(x) = g(a, b-1)$ . On the other hand, deleting the final east step from a path in  $L_1$  defines a bijection from  $L_1$  to  $L(a-1, b)$  that is *not* weight-preserving. The reason is that the  $b$  area cells below the final east step in a path in  $L_1$  no longer contribute to the area of the path in  $L(a-1, b)$ . However, since the area drops by  $b$  for all objects in  $L_1$ , we can conclude that  $G_{L_1, \text{area}}(x) = x^b G_{L(a-1, b), \text{area}}(x) = x^b g(a-1, b)$ . Now, by the sum rule for weighted sets,

$$g(a, b) = G_{L(a, b), \text{area}}(x) = G_{L_1}(x) + G_{L_2}(x) = x^b g(a-1, b) + g(a, b-1).$$

We remark that a similar argument involving deletion of the initial step of a path in  $L(a, b)$  establishes the dual recursion  $g(a, b) = g(a-1, b) + x^a g(a, b-1)$ .

*Second Proof.* Here we will prove that

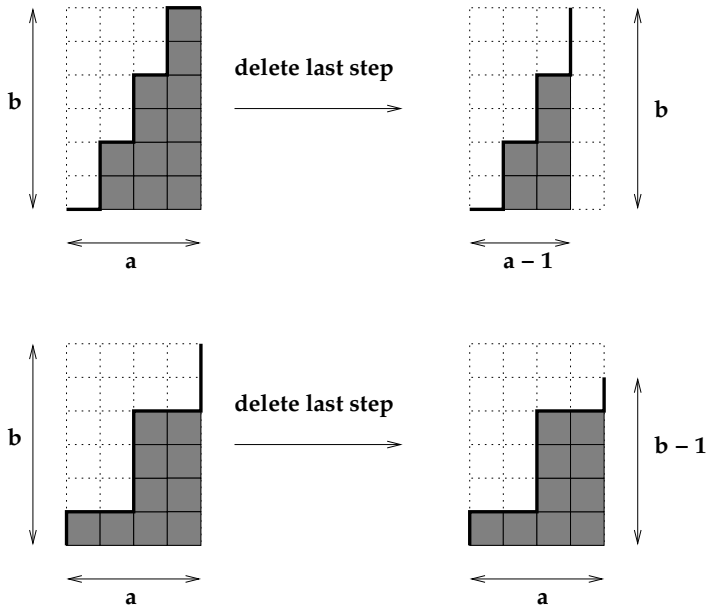
$$[a+b]!_x = [a]!_x [b]!_x \sum_{w \in \mathcal{R}(0^a 1^b)} x^{\text{inv}(w)},$$

which is equivalent to the first equality in the theorem statement. We know from 6.26 that the left side here is the generating function for the set  $S_{a+b}$  of permutations of  $\{1, 2, \dots, a+b\}$ , weighted by inversions. By the product rule, the right side is the generating function for the product set  $S_a \times S_b \times \mathcal{R}(0^a 1^b)$ , with the usual product weight  $\text{wt}(u, v, w) = \text{inv}(u) + \text{inv}(v) + \text{inv}(w)$ . Therefore, it suffices to define a bijection

$$f : S_a \times S_b \times \mathcal{R}(0^a 1^b) \rightarrow S_{a+b}$$

such that  $\text{inv}(f(u, v, w)) = \text{inv}(u) + \text{inv}(v) + \text{inv}(w)$  for  $u \in S_a$ ,  $v \in S_b$ , and  $w \in \mathcal{R}(0^a 1^b)$ .



**FIGURE 6.2**

Deleting the final step of a lattice path.

Given  $(u, v, w)$  in the domain of  $f$ , note that  $u$  is a permutation of the  $a$  letters  $1, 2, \dots, a$ . Replace the  $a$  zeroes in  $w$  with these  $a$  letters, in the same order that they occur in  $u$ . Next, add  $a$  to each of the letters in the permutation  $v$ , and then replace the  $b$  ones in  $w$  by these new letters in the same order that they occur in  $v$ . The resulting object  $z$  is evidently a permutation of  $\{1, 2, \dots, a + b\}$ . For example, if  $a = 3$  and  $b = 5$ , then

$$f(132, 24531, 01100111) = 15732864.$$

Since  $a$  and  $b$  are fixed and known, we can reverse the action of  $f$ . Starting with a permutation  $z$  of  $a + b$  elements, we first recover the word  $w \in \mathcal{R}(0^a 1^b)$  by replacing the numbers  $1, 2, \dots, a$  in  $z$  by zeroes and replacing the numbers  $a + 1, \dots, a + b$  in  $z$  by ones. Next, we take the subword of  $z$  consisting of the numbers  $1, 2, \dots, a$  to recover  $u$ . Similarly, let  $v'$  be the subword of  $z$  consisting of the numbers  $a + 1, \dots, a + b$ . We recover  $v$  by subtracting  $a$  from each of these numbers. This algorithm defines a two-sided inverse map to  $f$ . For example, still taking  $a = 3$  and  $b = 5$ , we have

$$f^{-1}(35162847) = (312, 23514, 01010111).$$

All that remains is to check that  $f$  is weight-preserving. Fix  $u, v, w, z$  with  $z = f(u, v, w)$ . Let  $A$  be the set of positions in  $z$  occupied by letters in  $u$ , and let  $B$  be the remaining positions (occupied by shifted letters of  $v$ ). Equivalently, by definition of  $f$ ,  $A = \{i : w_i = 0\}$  and  $B = \{i : w_i = 1\}$ . The inversions of  $z$  can be classified into three kinds. First, there are inversions  $(i, j)$  such that  $i, j \in A$ . These inversions correspond bijectively to the inversions of  $u$ . Second, there are inversions  $(i, j)$  such that  $i, j \in B$ . These inversions correspond bijectively to the inversions of  $v$ . Third, there are inversions  $(i, j)$  such that  $i \in A$  and  $j \in B$ , or  $i \in B$  and  $j \in A$ . The first case ( $i \in A, j \in B$ ) cannot occur, because every position in  $A$  is filled with a lower number than every position in  $B$ . The second case ( $i \in B, j \in A$ ) occurs iff  $i < j$  and  $w_i = 1$  and  $w_j = 0$ . This means that the inversions of the

third kind in  $z$  correspond bijectively to the inversions of the binary word  $w$ . Conclusion:  $\text{inv}(z) = \text{inv}(u) + \text{inv}(v) + \text{inv}(w)$ , as desired.  $\square$

Like ordinary binomial coefficients, the quantum binomial coefficients appear in a plethora of identities, many of which have combinatorial proofs. Here is a typical example, which is a weighted analogue of the Chu-Vandermonde identity in 2.21.

**6.37. Theorem: Quantum Chu-Vandermonde Identity.** For all integers  $a, b, c \geq 0$ ,

$$\left[ \begin{matrix} a+b+c+1 \\ a, b+c+1 \end{matrix} \right]_x = \sum_{k=0}^a x^{(b+1)(a-k)} \left[ \begin{matrix} k+b \\ k, b \end{matrix} \right]_x \left[ \begin{matrix} a-k+c \\ a-k, c \end{matrix} \right]_x.$$

*Proof.* Recall the picture we used to prove the original version of the identity (Figure 2.3), which is reprinted here as Figure 6.3. The path dissection in this picture defines a bijection

$$f : L(a, b+c+1) \rightarrow \bigcup_{k=0}^a L(k, b) \times L(a-k, c).$$

Here,  $k$  is the  $x$ -coordinate where the given path in  $L(a, b+c+1)$  crosses the line  $y = b+(1/2)$ . The bijection  $f$  is not weight-preserving. However, if a path  $P \in L(a, b+c+1)$  maps to  $(Q, R) \in L(k, b) \times L(a-k, c)$  under  $f$ , then it is evident from the picture that

$$\text{area}(P) = \text{area}(Q) + \text{area}(R) + (b+1)(a-k).$$

(The extra factor comes from the lower-right rectangle of width  $a-k$  and height  $b+1$ .) It now follows from the weight-shifting rule, the sum rule, and the product rule that

$$G_{L(a, b+c+1), \text{area}}(x) = \sum_{k=0}^a x^{(b+1)(a-k)} G_{L(k, b), \text{area}}(x) \cdot G_{L(a-k, c), \text{area}}(x).$$

We complete the proof by using 6.36 to replace each area generating function here by a suitable quantum binomial coefficient.  $\square$

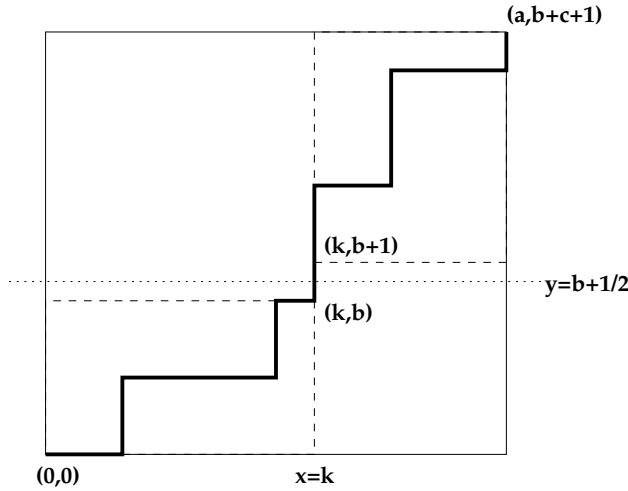
**6.38. Remark.** There are also linear-algebraic interpretations of the quantum binomial coefficients. Specifically, let  $F$  be a finite field with  $q$  elements, where  $q$  is necessarily a prime power. Then the integer  $\left[ \begin{matrix} n \\ k \end{matrix} \right]_q$  (which is  $\left[ \begin{matrix} n \\ k \end{matrix} \right]_x$  evaluated at  $x = q$ ) is the number of  $k$ -dimensional subspaces of the  $n$ -dimensional vector space  $F^n$ . We prove this fact in §12.7.

## 6.8 Quantum Multinomial Coefficients

Recall from 2.27 that the ordinary multinomial coefficients  $C(n; n_1, \dots, n_s)$  (where  $n = \sum_k n_k$ ) satisfy the recursion

$$C(n; n_1, \dots, n_k) = \sum_{k=1}^s C(n-1; n_1, \dots, n_k-1, \dots, n_s).$$

The quantum multinomial coefficients satisfy the following analogous recursion.

**FIGURE 6.3**

Picture used to prove the  $q$ -Chu-Vandermonde identity.

**6.39. Theorem: Recursions for Quantum Multinomial Coefficients.** Let  $n_1, \dots, n_s$  be nonnegative integers, and set  $n = \sum_{k=1}^s n_k$ . Then

$$\left[ \begin{matrix} n \\ n_1, \dots, n_s \end{matrix} \right]_x = \sum_{k=1}^s x^{n_1+n_2+\dots+n_{k-1}} \left[ \begin{matrix} n-1 \\ n_1, \dots, n_k-1, \dots, n_s \end{matrix} \right]_x.$$

(If  $n_k = 0$ , the  $k$ th summand on the right side is zero.) The initial condition is  $\left[ \begin{matrix} 0 \\ 0, \dots, 0 \end{matrix} \right]_x = 1$ . Moreover,  $\left[ \begin{matrix} n \\ n_1, \dots, n_s \end{matrix} \right]_x$  is a polynomial in  $x$  with coefficients in  $\mathbb{N}$ .

*Proof.* Neither side of the claimed recursion changes if we drop all  $n_i$ 's that are equal to zero; so, without loss of generality, assume every  $n_i$  is positive. We can create a common factor of  $[n-1]!_x / \prod_{j=1}^s [n_j]!_x$  on the right side by multiplying the  $k$ th summand by  $[n_k]_x / [n_k]_x$ , for  $1 \leq k \leq s$ . Pulling out this common factor, we are left with

$$\sum_{k=1}^s x^{n_1+n_2+\dots+n_{k-1}} [n_k]_x = \sum_{k=1}^s x^{n_1+\dots+n_{k-1}} (1 + x + x^2 + \dots + x^{n_k-1}).$$

The  $k$ th summand consists of the sum of consecutive powers of  $x$  starting at  $x^{n_1+\dots+n_{k-1}}$  and ending at  $x^{n_1+\dots+n_k-1}$ . Chaining these together, we see that the sum evaluates to  $x^0 + x^1 + \dots + x^{n-1} = [n]_x$ . Multiplying by the common factor mentioned above, we obtain

$$\frac{[n]!_x}{\prod_{j=1}^s [n_j]!_x} = \left[ \begin{matrix} n \\ n_1, \dots, n_s \end{matrix} \right]_x,$$

as desired. The initial condition is immediate. Finally, we deduce polynomiality of the quantum multinomial coefficients using induction on  $n$  and the recursion just proved, as in the proof of 6.34.  $\square$

**6.40. Theorem: Quantum Multinomial Coefficients and Inversions of Words.** Suppose  $A = \{a_1 < a_2 < \dots < a_s\}$  is a totally ordered alphabet. For all integers

$$n_1, \dots, n_s \geq 0,$$

$$\left[ \begin{matrix} n_1 + \dots + n_s \\ n_1, n_2, \dots, n_s \end{matrix} \right]_x = \sum_{w \in \mathcal{R}(a_1^{n_1} \dots a_s^{n_s})} x^{\text{inv}(w)}.$$

*Proof.* For all integers  $n_1, \dots, n_s$ , define

$$g(n_1, \dots, n_s) = \sum_{w \in \mathcal{R}(a_1^{n_1} \dots a_s^{n_s})} x^{\text{inv}(w)}.$$

(This is zero by convention if any  $n_i$  is negative.) By induction on  $\sum_k n_k$ , it suffices to show that  $g$  satisfies the recursion in 6.39. Now  $g(0, 0, \dots, 0) = x^0 = 1$ , so the initial condition is correct. Next, fix  $n_1, \dots, n_s \geq 0$ , and let  $W$  be the set of words appearing in the definition of  $g(n_1, \dots, n_s)$ . Write  $W$  as the disjoint union of sets  $W_1, \dots, W_s$ , where  $W_k$  consists of the words in  $W$  with first letter  $a_k$ . By the sum rule,

$$g(n_1, \dots, n_s) = G_W(x) = \sum_{k=1}^s G_{W_k}(x).$$

Fix a value of  $k$  in the range  $1 \leq k \leq s$  such that  $W_k$  is nonempty. Erasing the first letter of a word  $w$  in  $W_k$  defines a bijection from  $W_k$  to the set  $\mathcal{R}(a_1^{n_1} \dots a_k^{n_k-1} \dots a_s^{n_s})$ . The generating function for the latter set is  $g(n_1, \dots, n_k - 1, \dots, n_s)$ . The bijection in question does not preserve weights, because inversions involving the first letter of  $w \in W_k$  disappear when this letter is erased. However, no matter what word  $w$  we pick in  $W_k$ , the number of inversions that involve the first letter in  $w$  will always be the same. Specifically, this first letter (namely  $a_k$ ) will cause inversions with all of the  $a_1$ 's,  $a_2$ 's,  $\dots$ , and  $a_{k-1}$ 's that follow it in  $w$ . The number of such letters is  $n_1 + \dots + n_{k-1}$ . Therefore, by the weight-shifting rule,

$$G_{W_k, \text{inv}}(x) = x^{n_1 + \dots + n_{k-1}} g(n_1, \dots, n_k - 1, \dots, n_s).$$

This equation is also correct if  $W_k = \emptyset$  (which occurs iff  $n_k = 0$ ). Using these results in the formula above, we conclude that

$$g(n_1, \dots, n_s) = \sum_{k=1}^s x^{n_1 + \dots + n_{k-1}} g(n_1, \dots, n_k - 1, \dots, n_s),$$

which is precisely the recursion occurring in 6.39. □

**6.41. Remark.** This theorem can also be proved by generalizing the second proof of 6.36. Specifically, one can prove that

$$[n_1 + \dots + n_s]!_x = [n_1]!_x \dots [n_s]!_x \sum_{w \in \mathcal{R}(1^{n_1} \dots s^{n_s})} x^{\text{inv}(w)}$$

by defining a weight-preserving bijection

$$f : S_{n_1 + \dots + n_s} \rightarrow S_{n_1} \times \dots \times S_{n_s} \times \mathcal{R}(1^{n_1} \dots s^{n_s})$$

(where  $S_{n_i}$  is the set of all permutations of  $\{1, 2, \dots, n_i\}$ , and all sets in the Cartesian product are weighted by inversions). We leave the details as an exercise.

## 6.9 Foata's Map

We know from 6.26 and 6.29 that  $\sum_{w \in S_n} x^{\text{inv}(w)} = [n]!_x = \sum_{w \in S_n} x^{\text{maj}(w)}$ , where  $S_n$  is the set of permutations of  $\{1, 2, \dots, n\}$ . We can express this result by saying that the statistics  $\text{inv}$  and  $\text{maj}$  are *equidistributed* on  $S_n$ . We have just derived a formula for the distribution of  $\text{inv}$  on more general sets of words, namely

$$\sum_{w \in \mathcal{R}(1^{n_1} \dots s^{n_s})} x^{\text{inv}(w)} = \left[ \begin{matrix} n_1 + \dots + n_s \\ n_1, \dots, n_s \end{matrix} \right]_x.$$

Could it be true that  $\text{inv}$  and  $\text{maj}$  are still equidistributed on these more general sets? MacMahon [90] proved that this is indeed the case. We present a combinatorial proof of this result based on a bijection due to Dominique Foata. For each set  $S = \mathcal{R}(1^{n_1} \dots s^{n_s})$ , our goal is to define a weight-preserving bijection  $f : (S, \text{maj}) \rightarrow (S, \text{inv})$ .

To achieve our goal, let  $W$  be the set of *all* words in the alphabet  $\{1, 2, \dots, s\}$ . We shall define a function  $g : W \rightarrow W$  with the following properties: (a)  $g$  is a bijection; (b) for all  $w \in W$ ,  $w$  and  $g(w)$  are anagrams (§1.9); (c) if  $w$  is not the empty word, then  $w$  and  $g(w)$  have the same last letter; (d) for all  $w \in W$ ,  $\text{inv}(g(w)) = \text{maj}(w)$ . We can then obtain the desired weight-preserving bijections  $f$  by restricting  $g$  to the various anagram classes  $\mathcal{R}(1^{n_1} \dots s^{n_s})$ .

We will define  $g$  by recursion on the length of  $w \in W$ . If this length is 0 or 1, set  $g(w) = w$ . Then conditions (b), (c), and (d) hold in this case. Now suppose  $w$  has length  $n \geq 2$ . Write  $w = w'yz$ , where  $w' \in W$  and  $y, z$  are the last two letters of  $w$ . We can assume by induction that  $u = g(w'y)$  has already been defined, and that  $u$  is an anagram of  $w'y$  ending in  $y$  such that  $\text{inv}(u) = \text{maj}(w'y)$ . We will define  $g(w) = h_z(u)z$ , where  $h_z : W \rightarrow W$  is a certain map (to be described momentarily) that satisfies conditions (a) and (b) above. No matter what the details of the definition of  $h_z$ , it is already evident that  $g$  will satisfy conditions (b) and (c) for words of length  $n$ .

To motivate the definition of  $h_z$ , we first give a lemma that analyzes the effect on  $\text{inv}$  and  $\text{maj}$  of appending a letter to the end of a word. The lemma will use the following convenient notation. If  $u$  is any word and  $z$  is any letter, let  $n_{\leq z}(u)$  be the number of letters in  $u$  (counting repetitions) that are  $\leq z$ ; define  $n_{< z}(u)$ ,  $n_{> z}(u)$ , and  $n_{\geq z}(u)$  similarly.

**6.42. Lemma.** Suppose  $u$  is a word of length  $m$  with last letter  $y$ , and  $z$  is any letter.

- (a) If  $y \leq z$ , then  $\text{maj}(uz) = \text{maj}(u)$ . (b) If  $y > z$ , then  $\text{maj}(uz) = \text{maj}(u) + m$ .
- (c)  $\text{inv}(uz) = \text{inv}(u) + n_{> z}(u)$ .

*Proof.* All statements follow routinely from the definitions of  $\text{inv}$  and  $\text{maj}$ . □

Let us now describe the map  $h_z : W \rightarrow W$ . First,  $h_z$  sends the empty word to itself. Now suppose  $u$  is a nonempty word ending in  $y$ . There are two cases. **Case 1:**  $y \leq z$ . In this case, we break the word  $u$  into *runs* of consecutive letters such that the last letter in each run is  $\leq z$ , while all preceding letters in the run are  $> z$ . For example, if  $u = 1342434453552$  and  $z = 3$ , then the decomposition of  $u$  into runs is

$$u = 1/3/4, 2/4, 3/4, 4, 5, 3/5, 5, 2/$$

where we use slashes to delimit consecutive runs. Now,  $h_z$  operates on  $u$  by cyclically shifting the letters in each run one step to the right. Continuing the preceding example,

$$h_3(u) = 1/3/2, 4/3, 4/3, 4, 4, 5/2, 5, 5/.$$

What effect does this process have on  $\text{inv}(u)$ ? The last element in each run (which is  $\leq z$ ) is strictly less than all elements before it in its run (which are  $> z$ ). So, moving the last element to the front of its run causes the inversion number to drop by the number of elements  $> z$  in the run. Adding up these changes over all the runs, we see that

$$\text{inv}(h_z(u)) = \text{inv}(u) - n_{>z}(u) \quad (6.1)$$

in case 1. Furthermore, note that the first letter of  $h_z(u)$  is always  $\leq z$  in this case.

**Case 2:**  $y > z$ . Again we break the word  $u$  into runs, but here the last letter of each run must be  $> z$ , while all preceding letters in the run are  $\leq z$ . For example, if  $z = 3$  and  $u = 134243445355$ , we decompose  $u$  as

$$u = 1, 3, 4/2, 4/3, 4/4/5/3, 5/5/.$$

As before, we cyclically shift the letters in each run one step right, which gives

$$h_3(u) = 4, 1, 3/4, 2/4, 3/4/5/5, 3/5/$$

in our example. This time, the last element in each run is  $> z$  and is strictly greater than the elements  $\leq z$  that precede it in its run. So, the cyclic shift of each run will increase the inversion count by the number of elements  $\leq z$  in the run. Adding over all runs, we see that

$$\text{inv}(h_z(u)) = \text{inv}(u) + n_{\leq z}(u) \quad (6.2)$$

in case 2. Furthermore, note that the first letter of  $h_z(u)$  is always  $> z$  in this case.

In both cases,  $h_z(u)$  is an anagram of  $u$ . Moreover, we can invert the action of  $h_z$  as follows. Examination of the first letter of  $h_z(u)$  tells us whether we were in case 1 or case 2 above. To invert in case 1, break the word into runs whose first letter is  $\leq z$  and whose other letters are  $> z$ , and cyclically shift each run one step left. To invert in case 2, break the word into runs whose first letter is  $> z$  and whose other letters are  $\leq z$ , and cyclically shift each run one step left. We now see that  $h_z$  is a bijection. For example, to compute  $h_3^{-1}(1342434453552)$ , first write

$$1/3, 4/2, 4/3, 4, 4, 5/3, 5, 5/2/$$

and then cyclically shift to get the answer  $1/4, 3/4, 2/4, 4, 5, 3/5, 5, 3/2/$ .

Now we can return to the discussion of  $g$ . Recall that we have set  $g(w) = g(w'yz) = h_z(u)z$ , where  $u = g(w'y)$  is an anagram of  $w'y$  ending in  $y$  and satisfying  $\text{inv}(u) = \text{maj}(w'y)$ . To check condition (d) for this  $w$ , we must show that  $\text{inv}(h_z(u)z) = \text{maj}(w)$ . Again consider two cases. If  $y \leq z$ , then

$$\text{maj}(w) = \text{maj}(w'yz) = \text{maj}(w'y) = \text{inv}(u).$$

On the other hand, by the lemma and (6.1), we have

$$\text{inv}(h_z(u)z) = \text{inv}(h_z(u)) + n_{>z}(h_z(u)) = \text{inv}(u) - n_{>z}(u) + n_{>z}(u) = \text{inv}(u)$$

(observe that  $n_{>z}(h_z(u)) = n_{>z}(u)$  since  $h_z(u)$  and  $u$  are anagrams). In the second case, where  $y > z$ , we have

$$\text{maj}(w) = \text{maj}(w'yz) = \text{maj}(w'y) + n - 1 = \text{inv}(u) + n - 1.$$

On the other hand, the lemma and (6.2) give

$$\text{inv}(h_z(u)z) = \text{inv}(h_z(u)) + n_{>z}(h_z(u)) = \text{inv}(u) + n_{\leq z}(u) + n_{>z}(u) = \text{inv}(u) + n - 1,$$

		current word:
		2, 1, 3, 3, 1, 3, 2, 2
$h_1(2)$	$= 2;$	2, 1, 3, 3, 1, 3, 2, 2
$h_3(2, 1)$	$= 2, 1;$	2, 1, 3, 3, 1, 3, 2, 2
$h_3(2, 1, 3)$	$= 2, 1, 3;$	2, 1, 3, 3, 1, 3, 2, 2
$h_1(2, 1, 3, 3)$	$= 2, 3, 1, 3;$	2, 3, 1, 3, 1, 3, 2, 2
$h_3(2, 3, 1, 3, 1)$	$= 2, 3, 1, 3, 1;$	2, 3, 1, 3, 1, 3, 2, 2
$h_2(2, 3, 1, 3, 1, 3)$	$= 3, 2, 3, 1, 3, 1;$	3, 2, 3, 1, 3, 1, 2, 2
$h_2(3, 2, 3, 1, 3, 1, 2)$	$= 2, 3, 1, 3, 1, 3, 2;$	2, 3, 1, 3, 1, 3, 2, 2

**FIGURE 6.4**Computation of  $g(w)$ .

		current word:
		2, 1, 3, 3, 1, 3, 2, 2
$h_2^{-1}(2, 1, 3, 3, 1, 3, 2)$	$= 2, 3, 3, 1, 3, 1, 2;$	2, 3, 3, 1, 3, 1, 2, 2
$h_2^{-1}(2, 3, 3, 1, 3, 1)$	$= 3, 3, 2, 3, 1, 1;$	3, 3, 2, 3, 1, 1, 2, 2
$h_1^{-1}(3, 3, 2, 3, 1)$	$= 3, 3, 2, 1, 3;$	3, 3, 2, 1, 3, 1, 2, 2
$h_3^{-1}(3, 3, 2, 1)$	$= 3, 3, 2, 1;$	3, 3, 2, 1, 3, 1, 2, 2
$h_1^{-1}(3, 3, 2)$	$= 3, 3, 2;$	3, 3, 2, 1, 3, 1, 2, 2
$h_2^{-1}(3, 3)$	$= 3, 3;$	3, 3, 2, 1, 3, 1, 2, 2
$h_3^{-1}(3)$	$= 3;$	3, 3, 2, 1, 3, 1, 2, 2

**FIGURE 6.5**Computation of  $g^{-1}(w)$ .

since  $u$  has  $n - 1$  letters, each of which is either  $\leq z$  or  $> z$ .

It remains to prove that  $g$  is a bijection, by describing the two-sided inverse map  $g^{-1}$ . This is the identity map on words of length at most 1. To compute  $g^{-1}(uz)$ , first compute  $u' = h_z^{-1}(u)$ . Then return the answer  $g^{-1}(uz) = (g^{-1}(u'))z$ . Here is a nonrecursive description of the maps  $g$  and  $g^{-1}$ , obtained by “unrolling” the recursive applications of  $g$  and  $g^{-1}$  in the preceding definitions.

*To compute  $g(w_1w_2 \cdots w_n)$ :* for  $i = 2, \dots, n$  in this order, apply  $h_{w_i}$  to the first  $i - 1$  letters of the current word.

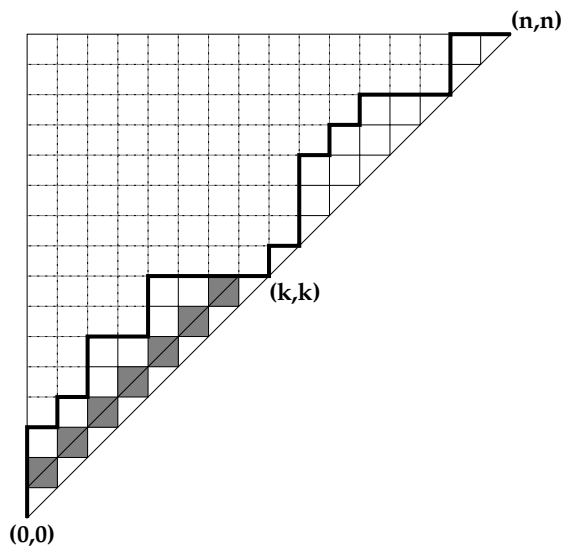
*To compute  $g^{-1}(z_1z_2 \cdots z_n)$ :* for  $i = n, n - 1, \dots, 2$  in this order, let  $z'_i$  be the  $i$ th letter of the current word, and apply  $h_{z'_i}^{-1}$  to the first  $i - 1$  letters of the current word.

**6.43. Example.** Figure 6.4 illustrates the computation of  $g(w)$  for  $w = 21331322$ . We find that  $g(w) = 23131322$ . Observe that  $\text{maj}(w) = 1 + 4 + 6 = 11 = \text{inv}(g(w))$ . Next, Figure 6.5 illustrates the calculation of  $g^{-1}(w)$ . We have  $g^{-1}(w) = 33213122$  and  $\text{inv}(w) = 10 = \text{maj}(g^{-1}(w))$ .

We summarize the results of this section in the following theorem.

**6.44. Theorem.** For all  $n_1, \dots, n_s \geq 0$ ,

$$\sum_{w \in \mathcal{R}(1^{n_1} \cdots s^{n_s})} x^{\text{maj}(w)} = \sum_{w \in \mathcal{R}(1^{n_1} \cdots s^{n_s})} x^{\text{inv}(w)} = \left[ \begin{matrix} n_1 + \cdots + n_s \\ n_1, \dots, n_s \end{matrix} \right]_x.$$

**FIGURE 6.6**

First-return analysis for weighted Dyck paths.

More precisely, there is a bijection on  $\mathcal{R}(1^{n_1} \cdots s^{n_s})$  sending  $\text{maj}$  to  $\text{inv}$  and preserving the last letter of each word.

---

## 6.10 Quantum Catalan Numbers

In this section, we investigate two weighted analogues of the Catalan numbers. Recall that the Catalan number  $C_n = \frac{1}{n+1} \binom{2n}{n} = \binom{2n}{n,n} - \binom{2n}{n-1,n+1}$  counts the collection of all lattice paths from  $(0,0)$  to  $(n,n)$  that never go below the line  $y = x$  (§1.10). Let  $D_n$  be the collection of these *Dyck paths*. Also, let  $W_n$  be the set of words that encode the Dyck paths, where we use 0 to encode a north step and 1 to encode an east step. Elements of  $W_n$  will be called *Dyck words*.

**6.45. Definition: Statistics on Dyck Paths.** For every Dyck path  $P \in D_n$ , let  $\text{area}(P)$  be the number of complete unit squares located between  $P$  and the line  $y = x$ . If  $P$  is encoded by the Dyck word  $w \in W_n$ , let  $\text{inv}(P) = \text{inv}(w)$  and  $\text{maj}(P) = \text{maj}(w)$ .

For example, the path  $P$  shown in Figure 6.6 has  $\text{area}(P) = 23$ . One sees that  $\text{inv}(P)$  is the number of unit squares in the region bounded by  $P$ , the  $y$ -axis, and the line  $y = n$ . We also have  $\text{inv}(P) + \text{area}(P) = \binom{n}{2}$  since  $\binom{n}{2}$  is the total number of area squares in the bounding triangle. The statistic  $\text{maj}(P)$  is the sum of the number of steps in the path that precede each “left-turn” where an east step (1) is immediately followed by a north step (0). For the path in Figure 6.6, we have  $\text{inv}(P) = 97$  and  $\text{maj}(P) = 4 + 6 + 10 + 16 + 18 + 22 + 24 + 28 = 128$ .



**6.46. Example.** When  $n = 3$ , examination of Figure 1.8 shows that

$$\begin{aligned} G_{D_3, \text{area}}(x) &= 1 + 2x + x^2 + x^3; \\ G_{D_3, \text{inv}}(x) &= 1 + x + 2x^2 + x^3; \\ G_{D_3, \text{maj}}(x) &= 1 + x^2 + x^3 + x^4 + x^6. \end{aligned}$$

When  $n = 4$ , a longer calculation gives

$$\begin{aligned} G_{D_4, \text{area}}(x) &= 1 + 3x + 3x^2 + 3x^3 + 2x^4 + x^5 + x^6; \\ G_{D_4, \text{maj}}(x) &= 1 + x^2 + x^3 + 2x^4 + x^5 + 2x^6 + x^7 + 2x^8 + x^9 + x^{10} + x^{12}. \end{aligned}$$

There is no particularly nice closed formula for  $G_{D_n, \text{area}}(x)$  (although determinant formulas do exist for this polynomial). However, these generating functions do satisfy a recursion, which is the analogue of the “first-return” recursion used in the unweighted case (§2.7).

**6.47. Theorem: Recursion for Dyck Paths Weighted by Area.** For all  $n \geq 0$ , set  $C_n(x) = G_{D_n, \text{area}}(x)$ . Then  $C_0(x) = 1$  and, for all  $n \geq 1$ ,

$$C_n(x) = \sum_{k=1}^n x^{k-1} C_{k-1}(x) C_{n-k}(x).$$

*Proof.* We imitate the proof of 2.33, but now we must take weights into account. For  $1 \leq k \leq n$ , write  $D_{n,k}$  for the set of Dyck paths of order  $n$  whose first return to the line  $y = x$  occurs at  $(k, k)$ . Evidently,  $D_n$  is the disjoint union of the  $D_{n,k}$ ’s, so the sum rule gives

$$C_n(x) = \sum_{k=1}^n G_{D_{n,k}, \text{area}}(x).$$

For fixed  $k$ , we have a bijection from  $D_{n,k}$  to  $D_{k-1} \times D_{n-k}$  defined by sending  $P = 0, P_1, 1, P_2$  to  $(P_1, P_2)$ , where the displayed 1 encodes the east step that arrives at  $(k, k)$ . See Figure 6.6. Examination of the figure shows that

$$\text{area}(P) = \text{area}(P_1) + \text{area}(P_2) + (k-1),$$

where the  $k-1$  counts the shaded cells in the figure which are not included in the calculation of  $\text{area}(P_1)$ . By the product rule and weight-shifting rule, we see that

$$G_{D_{n,k}, \text{area}}(x) = C_{k-1}(x) C_{n-k}(x) x^{k-1}.$$

Inserting this into the previous formula gives the recursion.  $\square$

Now let us consider the generating function  $G_{D_n, \text{maj}}(x)$ . This polynomial does have a nice closed formula, as we see in the next theorem.

**6.48. Theorem: Dyck Paths Weighted by Major Index.** For all  $n \geq 0$ ,

$$G_{D_n, \text{maj}}(x) = \left[ \begin{matrix} 2n \\ n, n \end{matrix} \right]_x - x \left[ \begin{matrix} 2n \\ n-1, n+1 \end{matrix} \right]_x = \frac{1}{[n+1]_x} \left[ \begin{matrix} 2n \\ n, n \end{matrix} \right]_x.$$

*Proof.* The second equality follows from the manipulation

$$\begin{aligned} \left[ \begin{matrix} 2n \\ n, n \end{matrix} \right]_x - x \left[ \begin{matrix} 2n \\ n-1, n+1 \end{matrix} \right]_x &= \left[ \begin{matrix} 2n \\ n, n \end{matrix} \right]_x \cdot \left( 1 - \frac{x[n]_x}{[n+1]_x} \right) \\ &= \left[ \begin{matrix} 2n \\ n, n \end{matrix} \right]_x \cdot \left( \frac{(1+x+x^2+\cdots+x^n) - x(1+x+\cdots+x^{n-1})}{[n+1]_x} \right) = \left[ \begin{matrix} 2n \\ n, n \end{matrix} \right]_x \cdot \frac{1}{[n+1]_x}. \end{aligned}$$

The other equality in the theorem statement can be rewritten

$$x \left[ \begin{matrix} 2n \\ n-1, n+1 \end{matrix} \right]_x + G_{D_n, \text{maj}}(x) = \left[ \begin{matrix} 2n \\ n, n \end{matrix} \right]_x.$$

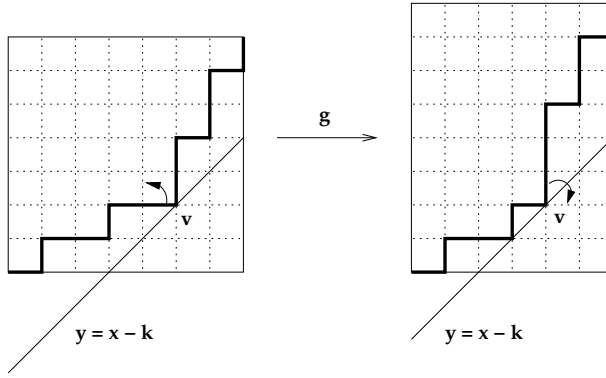
We will give a bijective proof of this result reminiscent of André's reflection principle (see 1.56). Consider the set of words  $S = \mathcal{R}(0^n 1^n)$ , weighted by major index. By 6.44, the generating function for this set is  $\left[ \begin{matrix} 2n \\ n, n \end{matrix} \right]_x$ . On the other hand, we can write  $S$  as the disjoint union of  $W_n$  and  $T$ , where  $W_n$  is the set of Dyck words and  $T$  consists of all other words in  $\mathcal{R}(0^n 1^n)$ . We will define a bijection  $g : T \rightarrow \mathcal{R}(0^{n+1} 1^{n-1})$  such that  $\text{maj}(w) = 1 + \text{maj}(g(w))$  for all  $w \in T$ . This will give

$$\left[ \begin{matrix} 2n \\ n, n \end{matrix} \right]_x = G_{S, \text{maj}}(x) = G_{T, \text{maj}}(x) + G_{W_n, \text{maj}}(x) = x \left[ \begin{matrix} 2n \\ n-1, n+1 \end{matrix} \right]_x + G_{D_n, \text{maj}}(x),$$

as desired. To define  $g(w)$  for  $w \in T$ , regard  $w$  as a lattice path in an  $n \times n$  rectangle by interpreting 0's as north steps and 1's as east steps. Find the largest  $k > 0$  such that the path  $w$  touches the line  $y = x - k$ . Such a  $k$  must exist, because  $w \in T$  is not a Dyck path. Consider the first vertex  $v$  on  $w$  that touches the line in question. (See Figure 6.7.) The path  $w$  must arrive at  $v$  by taking an east step, and  $w$  must leave  $v$  by taking a north step. These steps correspond to certain adjacent letters  $w_i = 1$  and  $w_{i+1} = 0$  in the word  $w$ . Furthermore, since  $v$  is the first visitation to this line, we must have either  $i = 1$  or  $w_{i-1} = 1$  (i.e., the step before  $w_i$  must be an east step if it exists). Let  $g(w)$  be the word obtained by changing  $w_i$  from 1 to 0. Pictorially, we “tip” the east step arriving at  $v$  upwards, changing it to a north step (which causes the following steps to shift to the northwest). The word  $w = \cdots 1, 1, 0 \cdots$  turns into  $g(w) = \cdots 1, 0, 0 \cdots$ , so the major index drops by exactly 1 when we pass from  $w$  to  $g(w)$ . This result also holds if  $i = 1$ . The new word  $g(w)$  has  $n - 1$  east steps and  $n + 1$  north steps, so  $g(w) \in \mathcal{R}(0^{n+1} 1^{n-1})$ . Finally,  $g$  is invertible. Given a path/word  $P \in \mathcal{R}(0^{n+1} 1^{n-1})$ , again take the largest  $k \geq 0$  such that  $P$  touches the line  $y = x - k$ , and let  $v$  be the *last* time  $P$  touches this line. Here,  $v$  is preceded by an east step and followed by two north steps (or  $v$  is the origin and is followed by a north step). Changing the first north step following  $v$  into an east step produces a path  $g'(P) \in \mathcal{R}(0^n 1^n)$ . One routinely checks that  $g'(P)$  cannot be a Dyck path (so  $g'$  maps into  $T$ ), and that  $g'$  is the two-sided inverse of  $g$ . The key is that the selection rules for  $v$  ensure that the same step is “tipped” when we apply  $g$  followed by  $g'$ , and similarly in the other order.  $\square$

## Summary

- *Generating Functions for Weighted Sets.* A weighted set is a pair  $(S, \text{wt})$  where  $S$  is a set and  $\text{wt} : S \rightarrow \mathbb{N}$  is a function (called a *statistic* on  $S$ ). The generating function for this weighted set is  $G_{S, \text{wt}}(x) = G_S(x) = \sum_{z \in S} x^{\text{wt}(z)}$ . Writing  $G_S(x) = \sum_{k \geq 0} a_k x^k$ ,  $a_k$  is the number of objects in  $S$  having weight  $k$ .
- *Weight-Preserving Bijections.* A weight-preserving bijection from  $(S, \text{wt})$  to  $(T, \text{wt}')$  is a bijection  $f : S \rightarrow T$  with  $\text{wt}'(f(z)) = \text{wt}(z)$  for all  $z \in S$ . When such an  $f$  exists,  $G_{S, \text{wt}}(x) = G_{T, \text{wt}'}(x)$ . More generally, if there is  $b \in \mathbb{Z}$  with  $\text{wt}'(f(z)) = b + \text{wt}(z)$  for all  $z \in S$ , then  $G_{T, \text{wt}'}(x) = x^b G_{S, \text{wt}}(x)$ .
- *Sum Rule for Weighted Sets.* Suppose  $(S_i, w_i)$  are weighted sets for  $1 \leq i \leq k$ ,  $S$  is the

**FIGURE 6.7**

The tipping bijection.

disjoint union of the  $S_i$ , and we define  $w : S \rightarrow \mathbb{N}$  by  $w(z) = w_i(z)$  for  $z \in S_i$ . Then  $G_S(x) = \sum_{i=1}^k G_{S_i}(x)$ .

- *Product Rule for Weighted Sets.* Suppose  $(S_i, w_i)$  are weighted sets for  $1 \leq i \leq k$ ,  $S = S_1 \times S_2 \times \cdots \times S_k$  is the product of the  $S_i$ , and we define  $w : S \rightarrow \mathbb{N}$  by  $w(z_1, \dots, z_k) = \sum_{i=1}^k w_i(z_i)$  for  $z_i \in S_i$ . Then  $G_S(x) = \prod_{i=1}^k G_{S_i}(x)$ .
- *Quantum Integers, Factorials, Binomial Coefficients, and Multinomial Coefficients.* Suppose  $x$  is a variable,  $n, k, n_i \in \mathbb{N}$ ,  $0 \leq k \leq n$ , and  $\sum_i n_i = n$ . We define  $[n]_x = \sum_{i=0}^{n-1} x^i = (x^n - 1)/(x - 1)$ ,  $[n]!_x = \prod_{i=1}^n [i]_x$ ,  $\begin{bmatrix} n \\ k \end{bmatrix}_x = \frac{[n]!_x}{[k]!_x [n-k]!_x}$ ,  $\begin{bmatrix} n \\ n_1, \dots, n_s \end{bmatrix}_x = \frac{[n]!_x}{\prod_{i=1}^s [n_i]!_x}$ . These are all polynomials in  $x$  with coefficients in  $\mathbb{N}$ .
- *Recursions for Quantum Binomial Coefficients, etc.* The following recursions hold:

$$\begin{aligned} [n]!_x &= [n-1]!_x \cdot [n]_x \\ \begin{bmatrix} a+b \\ a, b \end{bmatrix}_x &= x^b \begin{bmatrix} a+b-1 \\ a-1, b \end{bmatrix}_x + \begin{bmatrix} a+b-1 \\ a, b-1 \end{bmatrix}_x = \begin{bmatrix} a+b-1 \\ a-1, b \end{bmatrix}_x + x^a \begin{bmatrix} a+b-1 \\ a, b-1 \end{bmatrix}_x \\ \begin{bmatrix} n_1 + \cdots + n_s \\ n_1, \dots, n_s \end{bmatrix}_x &= \sum_{k=1}^s x^{n_1 + \cdots + n_{k-1}} \begin{bmatrix} n_1 + \cdots + n_s - 1 \\ n_1, \dots, n_k - 1, \dots, n_s \end{bmatrix}_x \end{aligned}$$

- *Statistics on Words.* Given a word  $w = w_1 w_2 \cdots w_n$  over a totally ordered alphabet,  $\text{Inv}(w) = \{(i, j) : i < j \text{ and } w_i > w_j\}$ ,  $\text{inv}(w) = |\text{Inv}(w)|$ ,  $\text{Des}(w) = \{i < n : w_i > w_{i+1}\}$ ,  $\text{des}(w) = |\text{Des}(w)|$ , and  $\text{maj}(w) = \sum_{i \in \text{Des}(w)} i$ . We have

$$\begin{bmatrix} n_1 + \cdots + n_s \\ n_1, \dots, n_s \end{bmatrix}_x = \sum_{w \in \mathcal{R}(a_1^{n_1} \cdots a_s^{n_s})} x^{\text{inv}(w)} = \sum_{w \in \mathcal{R}(a_1^{n_1} \cdots a_s^{n_s})} x^{\text{maj}(w)}.$$

The second equality follows from a subtle bijection due to Foata, which maps  $\text{maj}$  to  $\text{inv}$  while preserving the last letter of the word. In particular, letting  $S_n = \mathcal{R}(1^1 2^1 \cdots n^1)$ ,

$$[n]!_x = \sum_{w \in S_n} x^{\text{inv}(w)} = \sum_{w \in S_n} x^{\text{maj}(w)}.$$

These two formulas can be proved bijectively by mapping  $w \in S_n$  to its “inversion

table"  $(t_1, \dots, t_n)$ , where  $t_i$  records the change in inversions (resp. major index) caused by inserting the symbol  $i$  into the subword of  $w$  consisting of  $1, 2, \dots, i-1$ .

- *Weighted Lattice Paths.* The quantum binomial coefficient  $\left[ \begin{smallmatrix} a+b \\ a, b \end{smallmatrix} \right]_x = \left[ \begin{smallmatrix} b+a \\ b, a \end{smallmatrix} \right]_x$  counts lattice paths in an  $a \times b$  (or  $b \times a$ ) rectangle, weighted either by area above the path or area below the path. This coefficient also counts integer partitions with first part  $\leq a$  and length  $\leq b$ , weighted by area.
- *Weighted Dyck Paths.* Let  $C_n(x)$  be the generating function for Dyck paths of order  $n$ , weighted by area between the path and  $y = x$ . Then  $C_0(x) = 1$  and  $C_n(x) = \sum_{k=1}^n x^{k-1} C_{k-1}(x) C_{n-k}(x)$ . The generating function for Dyck paths (viewed as words in  $\mathcal{R}(0^n 1^n)$ ) weighted by major index is  $\frac{1}{[n+1]_x} \left[ \begin{smallmatrix} 2n \\ n, n \end{smallmatrix} \right]_x = \left[ \begin{smallmatrix} 2n \\ n, n \end{smallmatrix} \right]_x - x \left[ \begin{smallmatrix} 2n \\ n-1, n+1 \end{smallmatrix} \right]_x$ .

## Exercises

In the exercises below,  $S_n$  denotes the set  $\mathcal{R}(12 \cdots n)$  of permutations of  $\{1, 2, \dots, n\}$ , unless otherwise specified.

**6.49.** Let  $S = \mathcal{R}(a^1 b^1 c^2)$ ,  $T = \mathcal{R}(a^1 b^2 c^1)$ , and  $U = \mathcal{R}(a^2 b^1 c^1)$ . Confirm that  $G_{S, \text{inv}}(x) = G_{T, \text{inv}}(x) = G_{U, \text{inv}}(x)$  (as asserted in 6.15) by listing all weighted objects in  $T$  and  $U$ .

**6.50.** (a) Compute  $\text{inv}(w)$ ,  $\text{des}(w)$ , and  $\text{maj}(w)$  for each  $w \in S_4$ . (b) Use (a) to find the generating functions  $G_{S_4, \text{inv}}(x)$ ,  $G_{S_4, \text{des}}(x)$ , and  $G_{S_4, \text{maj}}(x)$ . (c) Compute  $[4]!_x$  by polynomial multiplication, and compare to the answers in (b).

**6.51.** (a) Compute  $\text{inv}(w)$  for the following words  $w$ : 4251673, 101101110001, 314423313, 55233514425331. (b) Compute  $\text{Des}(w)$ ,  $\text{des}(w)$ , and  $\text{maj}(w)$  for each word  $w$  in (a).

**6.52.** Confirm the formulas for  $G_{D_4, \text{area}}(x)$  and  $G_{D_4, \text{maj}}(x)$  stated in 6.46 by listing weighted Dyck paths of order 4.

**6.53.** (a) Find the maximum value of  $\text{inv}(w)$ ,  $\text{des}(w)$ , and  $\text{maj}(w)$  as  $w$  ranges over  $S_n$ . (b) Repeat (a) for  $w$  ranging over  $\mathcal{R}(1^{n_1} 2^{n_2} \cdots s^{n_s})$ .

**6.54.** Let  $S$  be the set of  $k$ -letter words over the alphabet  $\underline{n}$ . For  $w \in S$ , let  $\text{wt}(w)$  be the sum of all letters in  $w$ . Compute  $G_{S, \text{wt}}(x)$ .

**6.55.** Let  $S$  be the set of 5-letter words using the 26-letter English alphabet. For  $w \in S$ , let  $\text{wt}(w)$  be the number of vowels in  $w$ . Compute  $G_{S, \text{wt}}(x)$ .

**6.56.** Let  $S$  be the set of all subsets of  $\{1, 2, \dots, n\}$ . For  $A \in S$ , let  $\text{wt}(A) = |A|$ . Use the product rule for weighted sets to compute  $G_{S, \text{wt}}(x)$  (cf. 1.20).

**6.57.** Let  $S$  be the set of all  $k$ -element multisets using the alphabet  $\underline{n}$ . For  $M \in S$ , let  $\text{wt}(M)$  be the sum of the elements in  $M$ , counting multiplicities. Express  $G_{S, \text{wt}}(x)$  in terms of quantum binomial coefficients.

**6.58.** (a) How many permutations of  $\{1, 2, \dots, 8\}$  have exactly 17 inversions? (b) How many permutations of  $\{1, 2, \dots, 9\}$  have major index 29?

**6.59.** (a) How many lattice paths from  $(0, 0)$  to  $(8, 6)$  have area 21? (b) How many words in  $\mathcal{R}(0^5 1^6 2^7)$  have ten inversions? (c) How many Dyck paths of order 7 have major index 30?

**6.60. Quantum Binomial Theorem.** Let  $S$  be the set of all subsets of  $\{1, 2, \dots, n\}$ . For  $A \in S$ , let  $\text{wt}(A)$  be the sum of the elements in  $A$ . Show that

$$\prod_{i=1}^n (1 + x^i) = G_{S, \text{wt}}(x) = \sum_{k=0}^n x^{k(k+1)/2} \begin{bmatrix} n \\ k \end{bmatrix}_x.$$

**6.61.** Use an involution to prove  $\sum_{k=0}^n (-1)^k x^{k(k-1)/2} \begin{bmatrix} n \\ k \end{bmatrix}_x = 0$  for all  $n > 0$ .

**6.62.** Compute each of the following polynomials by any method, expressing the answer in the form  $\sum_{k \geq 0} a_k x^k$ : (a)  $[7]_x$ ; (b)  $[6]_x$ ; (c)  $\begin{bmatrix} 8 \\ 5 \end{bmatrix}_x$ ; (d)  $\begin{bmatrix} 7 \\ 2, 3, 2 \end{bmatrix}_x$ ; (e)  $\frac{1}{[8]_x} \begin{bmatrix} 8 \\ 5, 3 \end{bmatrix}_x$ ; (f)  $\frac{1}{[11]_x} \begin{bmatrix} 11 \\ 5, 6 \end{bmatrix}_x$ .

**6.63.** (a) Factor the polynomials  $[4]_x$ ,  $[5]_x$ ,  $[6]_x$ , and  $[12]_x$  in  $\mathbb{Z}[x]$ . (b) How do these polynomials factor in  $\mathbb{C}[x]$ ?

**6.64.** Compute  $\begin{bmatrix} 4 \\ 2 \end{bmatrix}_x$  in six ways, by: (a) simplifying the defining formula in 6.31; (b) using the first recursion in 6.33; (c) using the second recursion in 6.33; (d) enumerating words in  $\mathcal{R}(0011)$  by inversions; (e) enumerating words in  $\mathcal{R}(0011)$  by major index; (f) enumerating partitions contained in a  $2 \times 2$  box by area.

**6.65.** (a) Prove the identity  $\begin{bmatrix} n_1 + \dots + n_k \\ n_1, \dots, n_k \end{bmatrix}_x = \prod_{i=1}^k \begin{bmatrix} n_i + \dots + n_k \\ n_i, n_{i+1} + \dots + n_k \end{bmatrix}_x$  algebraically. (b) Give a combinatorial proof of the identity in (a).

**6.66.** For  $1 \leq i \leq 3$ , let  $(T_i, w_i)$  be a set of weighted objects. (a) Prove that  $\text{id}_{T_1} : T_1 \rightarrow T_1$  is a weight-preserving bijection. (b) Prove that if  $f : T_1 \rightarrow T_2$  is a weight-preserving bijection, then  $f^{-1} : T_2 \rightarrow T_1$  is weight-preserving. (c) Prove that if  $f : T_1 \rightarrow T_2$  and  $g : T_2 \rightarrow T_3$  are weight-preserving bijections, so is  $g \circ f$ .

**6.67.** Prove the second recursion in 6.33: (a) by an algebraic manipulation; (b) by removing the first step from a lattice path in an  $a \times b$  rectangle.

**6.68.** Let  $f$  be the map in the second proof of 6.36, with  $a = b = 4$ . Compute: (a)  $f(2413, 1423, 10011010)$ ; (b)  $f(4321, 4321, 11110000)$ ; (c)  $f(2134, 3214, 01010101)$ . Verify that weights are preserved in each case.

**6.69.** Let  $f$  be the map in the second proof of 6.36, with  $a = 5$  and  $b = 4$ . For each  $w$  given here, compute  $f^{-1}(w)$  and verify that weights are preserved: (a)  $w = 123456789$ ; (b)  $w = 371945826$ ; (c)  $w = 987456321$ .

**6.70.** Repeat 6.69 assuming  $a = 2$  and  $b = 7$ .

**6.71.** Prove that  $[n_1 + \dots + n_s]!_x = [n_1]!_x \dots [n_s]!_x \sum_{w \in \mathcal{R}(1^{n_1} \dots s^{n_s})} x^{\text{inv}(w)}$  by defining a weight-preserving bijection  $f : S_{n_1 + \dots + n_s} \rightarrow S_{n_1} \times \dots \times S_{n_s} \times \mathcal{R}(1^{n_1} \dots s^{n_s})$ .

**6.72.** (a) Find and prove an analogue of the identity  $\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}$  involving quantum binomial coefficients (cf. 2.19 and Figure 2.1). (b) Similarly, derive a quantum analogue of the identity  $\sum_{k=0}^a \binom{k+b-1}{k, b-1} = \binom{a+b}{a, b}$ .

**6.73.** Let  $S$  be the set of two-element subsets of Deck. For  $H \in S$ , let  $\text{wt}(H)$  be the sum of the values of the two cards in  $H$ , where aces count as 11 and jacks, queens, and kings count as 10. Find  $G_{S, \text{wt}}(x)$ .

**6.74.** Define the weight of a five-card poker hand to be the number of face cards in the hand (the face cards are aces, jacks, queens, and kings). Compute the generating functions for the following sets of poker hands relative to this weight: (a) full house hands; (b) three-of-a-kind hands; (c) flush hands; (d) straight hands.

**6.75.** Define the weight of a five-card poker hand to be the number of diamond cards in the hand. Compute the generating functions for the following sets of poker hands relative to this weight: (a) full house hands; (b) three-of-a-kind hands; (c) flush hands; (d) straight hands.

**6.76.** Let  $T_n$  be the set of connected simple graphs with vertex set  $\{1, 2, \dots, n\}$ . Let the weight of a graph in  $T_n$  be the number of edges. Compute  $G_{T_n}(x)$  for  $1 \leq n \leq 5$ .

**6.77.** Let  $f_n$  and  $g_n$  be the maps in 6.26. Compute  $f_6(341265)$  and  $g_6(0, 0, 1, 3, 2, 3)$ , and verify that weights are preserved for these two objects.

**6.78.** Let  $f_n$  and  $g_n$  be the maps in 6.26. Compute  $f_8(35261784)$  and  $g_8(0, 1, 0, 3, 2, 4, 6, 5)$ , and verify that weights are preserved for these two objects.

**6.79.** In 6.26, we constructed an inversion table for  $w \in S_n$  by classifying inversions  $(i, j) \in \text{Inv}(w)$  based on the left-hand value  $w_i$ . Define a new map  $f : S_n \rightarrow \underline{1} \times \underline{2} \times \dots \times \underline{n}$  by classifying inversions  $(i, j) \in \text{Inv}(w)$  based on the right-hand value  $w_j$ . Show that  $f$  is a bijection, and compute  $f(35261784)$  and  $f^{-1}(0, 1, 0, 3, 2, 4, 6, 5)$ .

**6.80.** Define a map  $f : S_n \rightarrow \underline{1} \times \underline{2} \times \dots \times \underline{n}$  by setting  $f(w) = (t_n, \dots, t_1)$ , where  $t_i = |\{j : (i, j) \in \text{Inv}(w)\}|$ . Show that  $f$  is a bijection. (Informally,  $f$  classifies inversions of  $w$  based on the left position of the inversion pair.) Compute  $f(35261784)$  and  $f^{-1}(0, 1, 0, 3, 2, 4, 6, 5)$ .

**6.81.** Define a map  $f : S_n \rightarrow \underline{1} \times \underline{2} \times \dots \times \underline{n}$  that classifies inversions of  $w$  based on the right position of the inversion pair (cf. 6.80). Show that  $f$  is a bijection, and compute  $f(35261784)$  and  $f^{-1}(0, 1, 0, 3, 2, 4, 6, 5)$ .

**6.82.** Let  $f_n$  be the map in 6.29. Compute  $f_6(341265)$  and  $f_6^{-1}(0, 0, 1, 3, 2, 3)$ , and verify that weights are preserved for these two objects.

**6.83.** Let  $f_n$  be the map in 6.29. Compute  $f_8(35261784)$  and  $f_8^{-1}(0, 1, 0, 3, 2, 4, 6, 5)$ , and verify that weights are preserved for these two objects.

**6.84. Coinversions.** Define the *coinversions* of a word  $w = w_1 w_2 \dots w_n$  by  $\text{coinv}(w) = \sum_{i < j} \chi(w_i < w_j)$ . Prove that  $\sum_{w \in \mathcal{R}(1^{n_1} 2^{n_2} \dots s^{n_s})} x^{\text{coinv}(w)} = \left[ \begin{smallmatrix} n_1 + \dots + n_s \\ n_1, \dots, n_s \end{smallmatrix} \right]_x$  (a) by using a bijection to reduce to the corresponding result for  $\text{inv}$ ; (b) by verifying a suitable recursion.

**6.85.** Given a word  $w = w_1 \dots w_n$ , define  $\text{comaj}(w) = \sum_{i < n} i \chi(w_i < w_{i+1})$  and  $\text{rlmaj}(w) = \sum_{i < n} (n - i) \chi(w_i > w_{i+1})$ . Calculate  $\sum_{w \in S_n} x^{\text{comaj}(w)}$  and  $\sum_{w \in S_n} x^{\text{rlmaj}(w)}$ .

**6.86.** For  $w \in S_n$ , let  $\text{wt}(w)$  be the sum of all  $i < n$  such that  $i + 1$  appears to the left of  $i$  in  $w$ . Compute  $G_{S_n, \text{wt}}(x)$ .

**6.87.** (a) Suppose  $w = w_1 w_2 \dots w_{n-1}$  is a fixed permutation of  $n - 1$  distinct letters. Let  $a$  be a new letter less than all letters appearing in  $w$ . Let  $S$  be the set of  $n$  words that can be obtained from  $w$  by inserting  $a$  in some gap. Prove that  $\sum_{z \in S} x^{\text{maj}(z)} = x^{\text{maj}(w)} [n]_x$ . (b) Use (a) to obtain another proof that  $\sum_{w \in S_n} x^{\text{maj}(w)} = [n]_x!$ .

**6.88.** Suppose  $k$  is fixed in  $\{1, 2, \dots, n\}$ , and  $w = w_1 w_2 \dots w_{n-1}$  is a fixed permutation of  $\{1, 2, \dots, k - 1, k + 1, \dots, n\}$ . Let  $S$  be the set of  $n$  words that can be obtained from  $w$  by inserting  $k$  in some gap. Prove or disprove:  $\sum_{z \in S} x^{\text{maj}(z)} = x^{\text{maj}(w)} [n]_x$ .

**6.89.** Define a *cyclic shift* function  $c : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  by  $c(i) = i + 1$  for  $i < n$ , and  $c(n) = 1$ . Define a map  $C : S_n \rightarrow S_n$  by setting  $C(w_1 w_2 \dots w_n) = c(w_1) c(w_2) \dots c(w_n)$ . (a) Prove: for all  $w \in S_n$ ,  $\text{maj}(C(w)) = \text{maj}(w) - 1$  if  $w_n \neq n$ , and  $\text{maj}(C(w)) = \text{maj}(w) + n - 1$  if  $w_n = n$ . (b) Use (a) to show combinatorially that, for  $1 \leq k \leq n$ ,  $\sum_{w \in S_n: w_n = k} x^{\text{maj}(w)} = x^{n-k} \sum_{v \in S_{n-1}} x^{\text{maj}(v)}$ . (c) Use (b), the sum rule, and induction to obtain another proof of 6.29.

**6.90.** For all  $n \geq 1$ , all  $T \subseteq \{1, 2, \dots, n-1\}$ , and  $1 \leq k \leq n$ , let  $G(n, T, k)$  be the number of permutations  $w$  of  $\{1, 2, \dots, n\}$  with  $\text{Des}(w) = T$  and  $w_n = k$ . (a) Find a recursion for the quantities  $G(n, T, k)$ . (b) Count the number of permutations of 10 objects with descent set  $\{2, 3, 5, 7\}$ .

**6.91.** Let  $w$  be the word 4523351452511332, and let  $h_z$  be the map from §6.9. Compute  $h_z(w)$  for  $z = 1, 2, 3, 4, 5, 6$ . Verify that (6.1) or (6.2) holds in each case.

**6.92.** Let  $w$  be the word 4523351452511332, and let  $h_z$  be the map from §6.9. Compute  $h_z^{-1}(w)$  for  $z = 1, 2, 3, 4, 5, 6$ .

**6.93.** Compute the image of each  $w \in S_4$  under the map  $g$  from §6.9.

**6.94.** Let  $g$  be the map in §6.9. Compute  $g(w)$  for each of these words: (a) 4251673; (b) 27418563; (c) 101101110001; (d) 314423313. Verify that  $\text{inv}(g(w)) = \text{maj}(w)$  in each case.

**6.95.** Let  $g$  be the map in §6.9. Compute  $g^{-1}(w)$  for each word  $w$  in 6.94.

**6.96.** Let  $g$  be the bijection in the proof of 6.48. Compute  $g(w)$  for each non-Dyck word  $w \in \mathcal{R}(0^3 1^3)$ .

**6.97. Quantum Fibonacci Numbers.** (a) Let  $W_n$  be the set of words in  $\{0, 1\}^n$  with no two consecutive zeroes, and let the weight of a word be the number of zeroes in it. Find a recursion for the generating functions  $G_{W_n}(x)$ , and use this to compute  $G_{W_6}(x)$ . (b) Repeat part (a), taking the weight to be the number of ones in the word.

**6.98.** Let  $S_{n,k}$  be the set of non-attacking placements of  $n-k$  rooks on the board  $\Delta_n$  (see 2.63). Define the weight of such a placement as follows. Each rook in the placement “cancels” all squares above it in its column. The weight of the placement is the total number of uncanceled squares located due west of rooks in the placement. Find a recursion for the generating functions  $G_{n,k} = G_{S_{n,k}, \text{wt}}(x)$ , which are quantum analogues of the Stirling numbers of the second kind. Compute these generating functions for  $0 \leq k \leq n \leq 5$ .

**6.99.** Let  $C_{n,k}$  be the set of permutations of  $\underline{n}$  consisting of  $k$  disjoint cycles. Define a statistic on permutations  $w \in C_{n,k}$  so that the associated generating functions satisfy the recursion

$$G_{C_{n,k}}(x) = G_{C_{n-1,k-1}}(x) + [n-1]_x G_{C_{n-1,k}}(x).$$

**6.100.** Let  $T_n$  be the set of trees with vertex set  $\underline{n}$ . Can you find a statistic on trees such that the associated generating function satisfies  $G_{T_n}(x) = [n]_x^{n-2}$ ?

**6.101. Multivariable Generating Functions.** Suppose  $S$  is a finite set, and  $w_1, \dots, w_n : S \rightarrow \mathbb{N}$  are  $n$  statistics on  $S$ . The *generating function for  $S$  relative to the  $n$  weights  $w_1, \dots, w_n$*  is the polynomial  $G_{S, w_1, \dots, w_n}(x_1, \dots, x_n) = \sum_{z \in S} \prod_{i=1}^n x_i^{w_i(z)}$ . Formulate and prove versions of the sum rule, product rule, bijection rule, and weight-shifting rule for such generating functions.

**6.102.** Extend 6.60 to a formula for  $\prod_{i=1}^n (1 + tx^i)$  by weighting subsets of  $\{1, 2, \dots, n\}$  by the number of elements in the subset and by the sum of the elements in the subset.

**6.103.** Recall from 1.29 that we can view permutations  $w \in S_n$  as bijective maps of  $\{1, 2, \dots, n\}$  into itself. Define  $I : S_n \rightarrow S_n$  by  $I(w) = w^{-1}$  for  $w \in S_n$ . (a) Show that  $I \circ I = \text{id}_{S_n}$ . (b) Show that  $\text{inv}(I(w)) = \text{inv}(w)$  for all  $w \in S_n$ . (c) Define  $\text{imaj}(w) = \text{maj}(I(w))$  for all  $w \in S_n$ . Compute the two-variable generating function  $G_n(x, y) = \sum_{w \in S_n} x^{\text{maj}(w)} y^{\text{imaj}(w)}$  for  $1 \leq n \leq 4$ . Prove that  $G_n(x, y) = G_n(y, x)$ .

**6.104.** Let  $g$  be the map in §6.9, and let  $\text{IDes}(w) = \text{Des}(w^{-1})$  for  $w \in S_n$ . (a) Show that for all  $w \in S_n$ ,  $\text{IDes}(g(w)) = \text{IDes}(w)$ . (b) Construct a bijection  $h : S_n \rightarrow S_n$  such that, for all  $w \in S_n$ ,  $\text{inv}(h(w)) = \text{maj}(w)$  and  $\text{maj}(h(w)) = \text{inv}(w)$ .

**6.105.** Let  $P_n$  be the set of integer partitions whose diagrams fit in the diagram of  $(n-1, n-2, \dots, 2, 1, 0)$ , i.e.,  $\mu \in P_n$  iff  $\ell(\mu) < n$  and  $\mu_i \leq n-i$  for  $1 \leq i < n$ . Let  $G_n(x) = \sum_{\mu \in P_n} x^{|\mu|}$ . Find a recursion satisfied by  $G_n(x)$  and use this to calculate  $G_5(x)$ . What is the relation between  $G_n(x)$  and the quantum Catalan number  $C_n(x)$  from §6.10?

**6.106. Bounce Statistic on Dyck Paths.** Given a Dyck path  $P \in D_n$ , define a new weight  $\text{bounce}(P)$  as follows. A ball starts at  $(0, 0)$  and moves north and east to  $(n, n)$  according to the following rules. The ball moves north  $v_0$  steps until blocked by an east step of  $P$ , then moves east  $v_0$  steps to the line  $y = x$ . The ball then moves north  $v_1$  steps until blocked by the east step of  $P$  starting on the line  $x = v_0$ , then moves east  $v_1$  steps to the line  $y = x$ . This bouncing process continues, generating a sequence  $(v_0, v_1, \dots, v_s)$  of vertical moves adding to  $n$ . We define  $\text{bounce}(P) = \sum_{i=0}^s i v_i$  and  $C_n(q, t) = \sum_{P \in D_n} q^{\text{area}(P)} t^{\text{bounce}(P)}$ . (a) Calculate  $C_n(q, t)$  for  $n \leq 4$  by enumerating Dyck paths. (b) Let  $C_{n,k}(q, t) = \sum_{P \in D_n: v_0(P)=k} q^{\text{area}(P)} t^{\text{bounce}(P)}$  be the generating function for Dyck paths that start with exactly  $k$  north steps. Establish the recursion

$$C_{n,k}(q, t) = \sum_{r=1}^{n-k} t^{n-k} q^{k(k-1)/2} \begin{bmatrix} r+k-1 \\ r, k-1 \end{bmatrix}_q C_{n-k,r}(q, t)$$

by “removing the first bounce.” Show also that  $C_n(q, t) = t^{-n} C_{n+1,1}(q, t)$ . (c) Use the recursion in (b) to calculate  $C_n(q, t)$  for  $n = 5, 6$ . (d) Prove that  $q^{n(n-1)/2} C_n(q, 1/q) = \sum_{P \in D_n} q^{\text{maj}(P)}$ . (e) Can you prove  $C_n(q, t) = C_n(t, q)$  for all  $n \geq 1$ ?

**6.107.** Let  $G_n$  be the set of sequences  $g = (g_0, g_1, \dots, g_{n-1})$  of nonnegative integers with  $g_0 = 0$  and  $g_{i+1} \leq g_i + 1$  for all  $i < n-1$  (cf. 2.120). For  $g \in G_n$ , define  $\text{area}(g) = \sum_{i=0}^{n-1} g_i$  and  $\text{dinv}(g) = \sum_{i < j} \chi(g_i - g_j \in \{0, 1\})$ . (a) Find a bijection  $k : G_n \rightarrow D_n$  such that  $\text{area}(k(g)) = \text{area}(g)$  for all  $g \in G_n$ . (b) Find a bijection  $h : G_n \rightarrow D_n$  such that  $\text{area}(h(g)) = \text{dinv}(g)$  and  $\text{bounce}(h(g)) = \text{area}(g)$  for all  $g \in G_n$  (see 6.106). Conclude that the statistics  $\text{dinv}$ ,  $\text{bounce}$ ,  $\text{area}$  (on  $G_n$ ), and  $\text{area}$  (on  $D_n$ ) all have the same distribution.

## Notes

The idea used to prove 6.29 seems to have first appeared in Gupta [63]. The bijection in §6.9 is due to Foata [38]. For related material, see Foata and Schützenberger [40]. Much of the early work on permutation statistics, including proofs of 6.44 and 6.48, is due to Major Percy MacMahon [90]. The bijective proof of 6.48, along with other material on quantum Catalan numbers, may be found in Fülringler and Hofbauer [47]. The bounce statistic in 6.106 was introduced by Haglund [64]; for more on this topic, see Haglund [65].



This page intentionally left blank

## Formal Power Series

In the last chapter, we introduced techniques for computing generating functions  $G_S(x) = \sum_{z \in S} x^{\text{wt}(z)}$  where  $S$  is a *finite* set of weighted objects. These generating functions are polynomials in the variable  $x$ . Now suppose that  $S$  is an *infinite* set of weighted objects. By analogy with the finite case, we would like to define a generating function  $G_S(x) = \sum_{z \in S} x^{\text{wt}(z)} = \sum_{n \geq 0} a_n x^n$ , where  $a_n$  is the number of objects in  $S$  of weight  $n$ . But the resulting expression  $G_S(x)$  is no longer a polynomial in  $x$ , since a polynomial can have only finitely many terms.

For example, if  $S$  is the set of all words over the alphabet  $\{0, 1\}$  weighted by length, we have  $a_n = 2^n$  for all  $n \geq 0$ , and so

$$G_S(x) = 1 + 2x + 4x^2 + 8x^3 + \cdots + 2^n x^n + \cdots = \sum_{n=0}^{\infty} 2^n x^n.$$

If we think of  $x$  as a real number, then  $G_S(x)$  is a *function of a real variable* that is defined for all  $x$  sufficiently close to zero. In fact, using the geometric series formula from calculus, one sees that  $G_S(x) = \sum_{n \geq 0} (2x)^n = \frac{1}{1-2x}$  for  $-1/2 < x < 1/2$ . For values of  $x$  outside this interval,  $G_S(x)$  is undefined. More generally, the power series  $\sum_{n=0}^{\infty} a_n x^n$  can be regarded as a function of a real (or complex) variable  $x$  that is defined within a certain interval of convergence centered at  $x = 0$ . However, difficulties can emerge if the coefficients  $a_n$  grow too rapidly. For instance, given  $a_n = n!$  for all  $n$ , one can show using the ratio test that the power series  $H(x) = \sum_{n=0}^{\infty} n! x^n$  only converges at  $x = 0$ . Thus we cannot recover the coefficients  $a_n = n!$  from knowledge of the function  $H$ , which is only defined at  $x = 0$ .

As this example shows, using real-valued functions to model combinatorial generating functions can be problematic because one must constantly worry about questions of convergence. We would prefer a purely *formal* notion of a power series in which convergence issues do not arise. The idea is to view a generating function  $\sum_{n=0}^{\infty} a_n x^n$  as merely a shorthand for an infinite sequence of integers  $(a_0, a_1, a_2, \dots, a_n, \dots)$ . The letter  $x$  is now only a symbol, not a variable; we are *not* allowed to substitute specific real numbers for  $x$ .

This chapter gives a rigorous development of the algebraic properties of formal power series. Our goal is to extend the familiar operations on polynomial functions (like addition, multiplication, composition, and differentiation) to the setting of formal power series. In certain situations, we will even be able to define infinite sums and products of formal power series. These algebraic operations will be used to help develop the combinatorics of infinite weighted sets, which is the topic of the next chapter.

In combinatorics, it usually suffices to consider formal power series whose coefficients are integers, rational numbers, or complex numbers. In this chapter, we will consider the slightly more general situation where the coefficients come from any field of characteristic zero (cf. 7.1 below). In fact, much of the algebraic theory is valid for power series with coefficients coming from an arbitrary ring (see 2.2). We shall indicate, as we proceed, which proofs require the stronger assumptions we are imposing on the coefficient ring.

## 7.1 The Ring of Formal Power Series

**7.1. Notational Convention.** Throughout this chapter, the letter  $K$  will stand for a field (see 2.3) that contains the field  $\mathbb{Q}$  of rational numbers.

For example,  $K$  might be  $\mathbb{Q}$  itself, or  $\mathbb{R}$  (the field of real numbers), or  $\mathbb{C}$  (the field of complex numbers).  $K$  might also be a field  $\mathbb{Q}(x)$  of formal rational functions, discussed in 7.46 below.

**7.2. Definition: Formal Power Series.** A *formal power series in one variable with coefficients in  $K$*  is a function  $F : \mathbb{N} \rightarrow K$ . We write  $F(n)$  or  $F_n$  for the value of the function  $F$  on the input  $n \in \mathbb{N}$ . The set of all such functions will be denoted  $K[[x]]$ , where  $x$  is a symbol called an *indeterminate*.

A formal power series  $F \in K[[x]]$  is exactly the same as a *sequence*

$$F = (F_0, F_1, F_2, \dots, F_n, \dots) = (F(0), F(1), F(2), \dots, F(n), \dots)$$

indexed by nonnegative integers, where each  $F_n \in K$ . We often display this sequence using *power series notation*, writing

$$F = \sum_{n=0}^{\infty} F_n x^n$$

and calling  $F_n$  the *coefficient of  $x^n$  in  $F$* . For the time being, the symbol  $x$  appearing in this notation has no independent meaning, and there are no addition, multiplication, or exponentiation operations being performed on the right side. This notation is used to help motivate the algebraic operations on power series to be introduced below, which are suggested by corresponding operations on one-variable polynomials.

**7.3. Remark: Equality of Formal Power Series.** Two formal power series  $F, G \in K[[x]]$  are *equal* iff  $F_n = G_n$  for all  $n \in \mathbb{N}$ . This follows from the definition of equality of two functions with domain  $\mathbb{N}$ .

**7.4. Example.** Consider the functions  $G, H : \mathbb{N} \rightarrow \mathbb{Q}$  defined by  $G(n) = 2^n$  and  $H(n) = n!$  for all  $n \in \mathbb{N}$ . These are two elements of  $\mathbb{Q}[[x]]$  which were discussed in the introduction to this chapter. In sequence notation and power series notation, we would write

$$G_S = (1, 2, 4, 8, \dots, 2^n, \dots) = \sum_{n \geq 0} 2^n x^n;$$

$$H = (1, 1, 2, 6, 24, 120, 720, \dots, n!, \dots) = \sum_{n \geq 0} n! x^n.$$

The function  $Z : \mathbb{N} \rightarrow K$  such that  $Z(n) = 0_K$  for all  $n \in \mathbb{N}$  defines a *zero power series*  $Z = \sum_{n \geq 0} 0x^n$ . We often denote  $Z$  (as well as the additive identity of  $K$ , the integer zero, etc.) by the symbol  $0$ .

**7.5. Example:  $X_i$ .** For each  $i \in \mathbb{N}$ , define a power series  $X_i : \mathbb{N} \rightarrow K$  by setting  $X_i(i) = 1$  and  $X_i(j) = 0$  for all  $j \neq i$ . Thus  $X_i$  is the sequence  $(0, 0, \dots, 1, 0, \dots)$  where the 1 is preceded by  $i$  zeroes. We have

$$X_i = \sum_{n=0}^{\infty} \chi(n=i) x^n.$$

If we omit zero coefficients, it is tempting to write  $X_0 = x^0 = 1$ ,  $X_1 = x^1 = x$ , and  $X_i = x^i$ . Strictly speaking, these abbreviations of the official power series notation are not allowed, but soon we will find a way to justify them.

We can now define addition and multiplication of formal power series.

**7.6. Definition: Sum and Product of Formal Power Series.** Given  $F, G \in K[[x]]$ , define the *sum*  $F + G : \mathbb{N} \rightarrow K$  by  $(F + G)(n) = F(n) + G(n)$  for all  $n \in \mathbb{N}$ . Define the *product*  $FG : \mathbb{N} \rightarrow K$  by

$$(FG)(n) = \sum_{\substack{(i,j) \in \mathbb{N}^2 \\ i+j=n}} F(i)G(j) = \sum_{k=0}^n F(k)G(n-k).$$

$FG$  is sometimes called the *convolution* of the functions  $F$  and  $G$ .

In sequence notation, this definition says

$$(F_n : n \geq 0) + (G_n : n \geq 0) = (F_n + G_n : n \geq 0);$$

$$(F_n : n \geq 0) \times (G_n : n \geq 0) = \left( \sum_{k=0}^n F_k G_{n-k} : n \geq 0 \right).$$

Using formal power series notation, these operations can also be written

$$\sum_{n \geq 0} F_n x^n + \sum_{n \geq 0} G_n x^n = \sum_{n \geq 0} (F_n + G_n) x^n;$$

$$\left( \sum_{n \geq 0} F_n x^n \right) \times \left( \sum_{n \geq 0} G_n x^n \right) = \sum_{n \geq 0} \left( \sum_{i+j=n} F_i G_j \right) x^n.$$

These formulas are exactly what we would expect (using the generalized distributive law) if  $x$  and every  $F_n$  and  $G_n$  were elements in some ring, and the sums appearing were finite.

**7.7. Example.** In  $\mathbb{Q}[[x]]$ , we have

$$(1, 2, 3, 4, 5, 6, \dots) + (1, 0, 1, 0, 1, 0, 1, 0, 1, 0, \dots) = (2, 2, 4, 4, 6, 6, \dots);$$

$$(1, 2, 3, 4, 5, 6, \dots) \times (1, 0, 1, 0, 1, 0, 1, 0, 1, 0, \dots) = (1, 2, 4, 6, 9, 12, 16, 20, \dots).$$

Given  $A = (3, 0, 2, 1, 7, 0, 0, 0, \dots)$  and  $B = (1, 4, 5, 0, 0, \dots)$  in  $K[[x]]$ , we have

$$A + B = (4, 4, 7, 1, 7, 0, 0, \dots);$$

$$AB = (3 \cdot 1, 3 \cdot 4 + 0 \cdot 1, 3 \cdot 5 + 0 \cdot 4 + 2 \cdot 1, \dots) = (3, 12, 17, 9, 21, 33, 35, 0, 0, \dots).$$

Compare these formal operations to the ordinary product of the two polynomial functions  $p(z) = 3 + 2z^2 + z^3 + 7z^4$  and  $q(z) = 1 + 4z + 5z^2$ :

$$p(z) + q(z) = 4 + 4z + 7z^2 + 1z^3 + 7z^4 \quad (z \in \mathbb{R});$$

$$p(z)q(z) = 3 + 12z + 17z^2 + 9z^3 + 21z^4 + 33z^5 + 35z^6 \quad (z \in \mathbb{R}).$$

Now suppose  $F = (F_n : n \in \mathbb{N}) \in K[[x]]$  and  $C = (1, 1, 1, \dots) \in K[[x]]$ . Then

$$FC = CF = (F_0, F_0 + F_1, F_0 + F_1 + F_2, \dots, F_0 + F_1 + \dots + F_n, \dots).$$

Thus, multiplication by  $C$  replaces a sequence of scalars by the sequence of partial sums of those scalars.

**7.8. Theorem: Algebraic Structure of  $K[[x]]$ .** (a) With the sum and product operations defined above,  $K[[x]]$  is a commutative ring. (b)  $K[[x]]$  contains the field  $K$ , provided we identify each  $a \in K$  with the sequence  $(a, 0, 0, 0, \dots) \in K[[x]]$ . (c)  $K[[x]]$  is a vector space over  $K$ .

*Proof.* (a) We verify some of the ring axioms for  $K[[x]]$  (see 2.2), leaving the others as exercises. If  $F$  and  $G$  are functions from  $\mathbb{N}$  to  $K$ ,  $F+G$  and  $FG$  are also functions that map  $\mathbb{N}$  into  $K$  (using closure of  $K$  under addition and multiplication). In other words,  $F \in K[[x]]$  and  $G \in K[[x]]$  imply  $F+G \in K[[x]]$  and  $FG \in K[[x]]$ , so the closure axioms for  $K[[x]]$  are true. To see that addition in  $K[[x]]$  is associative, fix  $F, G, H \in K[[x]]$ . Using associativity of addition in the field  $K$ , we see that

$$\begin{aligned} [(F+G)+H]_n &= (F+G)_n + H_n = (F_n + G_n) + H_n \\ &= F_n + (G_n + H_n) = F_n + (G+H)_n = [F+(G+H)]_n \end{aligned}$$

for every  $n \in \mathbb{N}$ . Thus (by 7.3)  $(F+G)+H = F+(G+H)$ .

The verification that  $(FG)H = F(GH)$  is somewhat more elaborate. On one hand, for a fixed  $n \in \mathbb{N}$ ,

$$[(FG)H]_n = \sum_{\substack{(i,c) \in \mathbb{N}^2: \\ i+c=n}} (FG)_i H_c = \sum_{\substack{(i,c) \in \mathbb{N}^2 \\ i+c=n}} \left( \sum_{\substack{(a,b) \in \mathbb{N}^2 \\ a+b=i}} (F_a G_b) \right) H_c = \sum_{\substack{(a,b,c) \in \mathbb{N}^3 \\ (a+b)+c=n}} (F_a G_b) H_c.$$

The last step used the distributive law in  $K$  and a reindexing of the summations (which is permissible since addition in  $K$  is commutative). On the other hand,

$$[F(GH)]_n = \sum_{\substack{(a,k) \in \mathbb{N}^2: \\ a+k=n}} F_a (GH)_k = \sum_{\substack{(a,k) \in \mathbb{N}^2 \\ a+k=n}} F_a \left( \sum_{\substack{(b,c) \in \mathbb{N}^2 \\ b+c=k}} (G_b H_c) \right) = \sum_{\substack{(a,b,c) \in \mathbb{N}^3 \\ a+(b+c)=n}} F_a (G_b H_c).$$

Using associativity of addition in  $\mathbb{N}$  and associativity of multiplication in  $K$ , we see that  $[(FG)H]_n = [F(GH)]_n$  for all  $n \in \mathbb{N}$ , hence  $(FG)H = F(GH)$  as desired.

Next we claim that  $X_0 = (1, 0, 0, \dots) = \sum_{n \geq 0} \chi(n=0)x^n$  is the multiplicative identity element in  $K[[x]]$ . For, given  $F \in K[[x]]$  and  $n \in \mathbb{N}$ , we compute

$$(X_0 F)_n = \sum_{i+j=n} X_0(i) F(j) = 1F(n) + 0F(n-1) + \dots + 0F(0) = F_n.$$

Thus  $X_0 F = F$ , and similarly  $F X_0 = F$ . We let the reader verify the remaining ring axioms, namely: the zero sequence is the additive identity in  $K[[x]]$ ; the additive inverse of  $(F_n : n \geq 0)$  is  $(-F_n : n \geq 0)$ ; addition in  $K[[x]]$  is commutative; multiplication in  $K[[x]]$  is commutative (this uses commutativity of  $K$ ); and the distributive law holds.

For (b), observe first that the map  $a \mapsto (a, 0, 0, \dots)$  is a bijection of  $K$  onto the subset of  $K[[x]]$  consisting of sequences that are zero after position zero. The definitions immediately show that

$$\begin{aligned} (a, 0, 0, \dots) + (b, 0, 0, \dots) &= (a+b, 0, 0, \dots); \\ -(a, 0, 0, \dots) &= (-a, 0, 0, \dots); \\ (a, 0, 0, \dots) \times (b, 0, 0, \dots) &= (ab, 0, 0, \dots); \end{aligned}$$

and furthermore,  $0_K \mapsto (0, 0, 0, \dots) = 0_{K[[x]]}$  and  $1_K \mapsto (1, 0, 0, \dots) = 1_{K[[x]]}$ . This shows

that operations in  $K[[x]]$  on sequences of this form agree with the corresponding field operations in  $K$ . So we can view  $K$  as embedded in  $K[[x]]$  by means of this bijection. (More formally, we have found an “isomorphic copy” of  $K$  inside  $K[[x]]$ .)

(c) Define scalar multiplication in  $K[[x]]$  by setting  $cF = (cF_n : n \in \mathbb{N})$  for  $c \in K$  and  $F \in K[[x]]$ . One checks that  $cF$  is the same as the product of  $(c, 0, 0, \dots)$  and  $F = (F_0, F_1, F_2, \dots)$  in the ring  $K[[x]]$ . Using this observation, one sees immediately that  $K[[x]]$  satisfies all the axioms for a vector space over  $K$ , because the required identities are special cases of the ring axioms that have just been verified.  $\square$

## 7.2 Finite Products and Powers of Formal Series

Now that we know  $K[[x]]$  is a ring, we can iterate the binary operations of addition and multiplication to define finite sums and products of formal power series. Similarly, for any integer  $n \geq 0$  and any  $G \in K[[x]]$ , the power  $G^n$  is defined recursively by setting  $G^0 = 1$  and, for  $n \geq 0$ , setting  $G^{n+1} = G^n \cdot G$ . Intuitively,  $G^n$  is the product of  $n$  factors all equal to  $G$ . Later, we will see that infinite sums and infinite products of formal power series can be defined in certain situations. We will also obtain a criterion for when the multiplicative inverse  $G^{-1}$  (and other negative powers of  $G$ ) can be formed.

**7.9. Example: Powers of  $x$ .** For  $i \in \mathbb{N}$ , define  $X_i = \sum_{n \geq 0} \chi(n=i)x^n$  as in 7.5. We claim that  $X_1^i = X_i$  for all  $i \geq 0$ . The claim holds when  $i = 0$  since  $X_1^0 = 1_{K[[x]]}$  by definition, and we saw in 7.8 that  $1_{K[[x]]} = X_0$ . Fix  $i \geq 0$  and assume by induction that  $X_1^i = X_i$ . Now

$$X_1^{i+1}(n) = (X_1^i \cdot X_1)(n) = (X_i \cdot X_1)(n) = \sum_{a+b=n} X_i(a)X_1(b) \quad (n \in \mathbb{N}).$$

The only choice of  $(a, b)$  that produces a nonzero summand is  $a = i$  and  $b = 1$ , which can occur only for  $n = i + 1$ . So  $X_1^{i+1}(n) = 0 = X_{i+1}(n)$  if  $n \neq i + 1$ , and  $X_1^{i+1}(i + 1) = 1 = X_{i+1}(i + 1)$ . Thus  $X_1^{i+1} = X_{i+1}$ , verifying the claim for  $i + 1$ .

If we *define*  $x$  to be the particular formal power series  $X_1 \in K[[x]]$ , the claim shows that  $x^i = X_i$  for all  $i \geq 0$ . We have now justified our earlier “notational abuse” in 7.5. Furthermore, for any *finite* sequence of scalars  $c_0, c_1, \dots, c_N \in K$ , define  $C \in K[[x]]$  by letting  $C(n) = c_n$  for  $0 \leq n \leq N$  and  $C(n) = 0$  for  $n > N$ . Then the definition of addition and scalar multiplication shows that

$$c_0 + c_1x + c_2x^2 + \cdots + c_Nx^N = (c_0, c_1, \dots, c_N, 0, 0, \dots) = C = \sum_{n \geq 0} C_n x^n,$$

where the leftmost expression is built up from  $x = X_1$  and the  $c_i$ ’s by algebraic operations in  $K[[x]]$ , and the rightmost expression is our atomic notation for the series  $C$ . Later, after we give a meaning to infinite summations of formal power series, we will see that the analogous identity

$$C_0 + C_1x + C_2x^2 + \cdots + C_nx^n + \cdots = C = \sum_{n \geq 0} C_n x^n$$

is also valid for any  $C \in K[[x]]$ .

**7.10. Theorem: Products of  $k$  Series.** Suppose  $G_1, G_2, \dots, G_k \in K[[x]]$ . For all  $n \in \mathbb{N}$ ,

$$(G_1 G_2 \cdots G_k)(n) = \sum_{\substack{(j_1, j_2, \dots, j_k) \in \mathbb{N}^k: \\ j_1 + j_2 + \cdots + j_k = n}} G_1(j_1) G_2(j_2) \cdots G_k(j_k).$$

*Proof.* We use induction on  $k$ . The case  $k = 1$  is immediate, and the case  $k = 2$  holds by the definition of the product of two formal power series. Assume  $k > 2$  and the result is already known for products of  $k - 1$  series. Letting  $F = G_1 G_2 \cdots G_{k-1}$ , we calculate

$$\begin{aligned}
 (G_1 G_2 \cdots G_k)(n) &= (F G_k)(n) = \sum_{\substack{(r,s) \in \mathbb{N}^2 \\ r+s=n}} F(r) G_k(s) \\
 &= \sum_{\substack{(r,s) \in \mathbb{N}^2 \\ r+s=n}} \left( \sum_{\substack{(j_1, j_2, \dots, j_{k-1}) \in \mathbb{N}^{k-1}: \\ j_1 + j_2 + \cdots + j_{k-1} = r}} G_1(j_1) G_2(j_2) \cdots G_{k-1}(j_{k-1}) \right) G_k(s) \\
 &= \sum_{\substack{(j_1, j_2, \dots, j_k) \in \mathbb{N}^k: \\ j_1 + j_2 + \cdots + j_k = n}} G_1(j_1) G_2(j_2) \cdots G_{k-1}(j_{k-1}) G_k(j_k).
 \end{aligned}$$

The last step follows by the generalized distributive law and a change in the names of the summation indices.  $\square$

Now we can prove a result, similar to the multinomial theorem 2.12, that lets us compute a power of a formal series.

**7.11. Theorem: Powers of Formal Series.** For all  $G \in K[[x]]$  and all  $m, n \in \mathbb{N}$ ,

$$G^m(n) = \sum_{\substack{(k_0, k_1, \dots, k_n) \in \mathbb{N}^{n+1}: \\ \sum_i k_i = m, \sum_i i k_i = n}} \binom{m}{k_0, k_1, \dots, k_n} G(0)^{k_0} G(1)^{k_1} \cdots G(n)^{k_n}.$$

*Proof.* Applying 7.10 with  $k = m$  and all  $G_i$ 's equal to  $G$ , we see that

$$G^m(n) = \sum_{\substack{(j_1, \dots, j_m) \in \mathbb{N}^m: \\ j_1 + \cdots + j_m = n}} G(j_1) G(j_2) \cdots G(j_m).$$

Given  $(k_0, k_1, \dots, k_n) \in \mathbb{N}^{n+1}$  satisfying  $\sum_i k_i = m$ ,  $\sum_i i k_i = n$ , let us group together all the summands indexed by sequences  $(j_1, \dots, j_m) \in \mathcal{R}(0^{k_0} 1^{k_1} \cdots n^{k_n})$ . The number of such summands is the multinomial coefficient  $\binom{m}{k_0, k_1, \dots, k_n}$  by 1.46, and (by commutativity of multiplication in  $K$ ) every such summand is equal to  $G(0)^{k_0} G(1)^{k_1} \cdots G(n)^{k_n}$ . Summing over all possible choices of  $(k_0, \dots, k_n)$  gives the stated formula for  $G^m(n)$ . (Compare to the proof of 2.12.)  $\square$

### 7.3 Formal Polynomials

Polynomials, like the power series studied in calculus, are often regarded as functions of a real variable  $x$ . For a function  $p : \mathbb{R} \rightarrow \mathbb{R}$  to be a polynomial, there must exist constants  $a_i \in \mathbb{R}$  and  $n \in \mathbb{N}$  such that  $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$  for all  $x \in \mathbb{R}$ . One can prove that the coefficients  $a_i$  are uniquely determined by  $p$ . This functional view of polynomials is not really necessary for many algebraic and combinatorial purposes. We now give a rigorous discussion of “formal” polynomials and their algebraic properties. Some of these properties will follow quickly from results already proved for formal power series.

**7.12. Definition: Formal Polynomials.** A formal power series  $F \in K[[x]]$  is a *polynomial* iff  $\{n \in \mathbb{N} : F(n) \neq 0\}$  is a finite set. Let  $K[x]$  be the set of all polynomials in  $K[[x]]$ .

Intuitively, a polynomial is a formal power series with only finitely many nonzero coefficients.

**7.13. Definition: Degree of a Polynomial.** Given a nonzero polynomial  $f \in K[x]$ , the *degree* of  $f$ , denoted  $\deg(f)$ , is the largest  $n \in \mathbb{N}$  with  $f(n) \neq 0$ . The element  $f(n) \in K$  is the *leading coefficient* of  $f$ . A polynomial  $f$  is called *monic* iff  $f(n) = 1$ . The degree of the zero polynomial is undefined.

**7.14. Theorem: Properties of Degree.** For all  $f, g \in K[x]$ , (a)  $f + g$  is a polynomial, and  $\deg(f + g) \leq \max(\deg(f), \deg(g))$  whenever both sides are defined. (b)  $fg$  is a polynomial. If  $f$  and  $g$  are nonzero, then  $fg \neq 0$ , and  $\deg(fg) = \deg(f) + \deg(g)$ .

*Proof.* (a) Certainly  $f + g$  is a polynomial if  $f = 0$  or  $g = 0$ . Otherwise, let  $n = \deg(f)$ ,  $m = \deg(g)$ , and  $k = \max(m, n)$ . For all  $i > k$ ,  $(f + g)(i) = f(i) + g(i) = 0 + 0 = 0$ . On one hand, this shows that  $\{i \in \mathbb{N} : (f + g)(i) \neq 0\} \subseteq \{0, 1, \dots, k\}$ , so that  $f + g$  is a polynomial. On the other hand, this also shows that  $\deg(f + g) \leq k$  if  $f + g \neq 0$ .

(b) Certainly  $fg$  is a polynomial if  $f = 0$  or  $g = 0$ . Now assume  $f$  is nonzero of degree  $n$ , and  $g$  is nonzero of degree  $m$ . Thus  $f(i) = 0$  for all  $i > n$  and  $g(j) = 0$  for all  $j > m$ . Suppose  $k > n + m$  and  $(i, j) \in \mathbb{N}^2$  satisfy  $i + j = k$ . Then we must have either  $i > n$  or  $j > m$ . Thus, every summand in the expression  $(fg)(k) = \sum_{i+j=k} f(i)g(j)$  is zero, so  $(fg)(k) = 0$  for all  $k > n + m$ . This shows that  $fg$  is a polynomial. We also have  $(fg)(n + m) = \sum_{i+j=n+m} f(i)g(j) = f(n)g(m)$  since the only nonzero summand occurs when  $i = n$  and  $j = m$ . Now  $f(n)g(m) \neq 0$  since  $f(n) \neq 0$  and  $g(m) \neq 0$  and  $K$  is a field (cf. 7.27 below). Thus,  $(fg)(n + m) \neq 0$ , whereas all higher coefficients of  $fg$  are zero. We now see that  $fg \neq 0$  and  $\deg(fg) = n + m = \deg(f) + \deg(g)$ .  $\square$

**7.15. Theorem: Algebraic Structure of  $K[x]$ .** (a)  $K[x]$  is a commutative ring containing the field  $K$ . (b)  $K[x]$  is a vector space over  $K$  with basis  $\{X_i : i \geq 0\} = \{x^i : i \geq 0\}$  (see 7.9).

*Proof.* (a) We have just seen that  $K[x]$  is closed under addition and multiplication. All the other ring axioms in 2.2 follow automatically from the corresponding ring axioms for  $K[[x]]$ , once we notice that  $-f$  is a polynomial whenever  $f$  is, and  $0_{K[[x]]}$  and  $1_{K[[x]]}$  are polynomials. More generally, for any  $c \in K$ , every power series of the form  $(c, 0, 0, \dots)$  is a polynomial. So  $K[x]$  contains the field  $K$  (or, more precisely, the copy of  $K$  inside  $K[[x]]$ ).

(b) Given a nonzero polynomial  $f \in K[x]$ , let  $n = \deg(f)$ . We see that

$$f(0)X_0 + f(1)X_1 + \cdots + f(n)X_n = f,$$

since both series take the same value at every  $k \in \mathbb{N}$  (cf. 7.9). Thus  $\{X_i : i \in \mathbb{N}\}$  is a spanning set for the vector space  $K[x]$ . To see that this set is linearly independent, consider a finite linear combination

$$c_{i_1}X_{i_1} + c_{i_2}X_{i_2} + \cdots + c_{i_k}X_{i_k} = 0,$$

where the  $i_j$ 's are distinct indices and each  $c_{i_j} \in K$ . Evaluating the left side at  $i_j$ , we see that  $c_{i_j} = 0$  for all  $j$ . Thus,  $\{X_i : i \geq 0\}$  is a basis of  $K[x]$ .  $\square$

Note that  $B = \{x^i : i \geq 0\}$  is a basis for  $K[x]$  but not a basis for  $K[[x]]$ . The set  $B$  does not span  $K[[x]]$ , because we are not allowed to form “infinite linear combinations”  $\sum_{i=0}^{\infty} c_i x^i$  when determining the span (in the linear-algebraic sense) of the set  $B$ . It is true that  $K[[x]]$  has some basis (as does every vector space) — but this basis will be much larger than the collection  $B$  and cannot be specified explicitly.



**7.16. Example.** The sequences  $f = (2, 0, 1, 3, 0, 0, \dots)$  and  $g = (1, -1, 0, -3, 0, 0, \dots)$  are polynomials of degree 3, which can be written in terms of the basis  $B$  as  $f = 2 + x^2 + 3x^3$  and  $g = 1 - x - 3x^3$ . We calculate

$$f + g = 3 - x + x^2, \quad fg = 2 - 2x + x^2 - 4x^3 - 3x^4 - 3x^5 - 9x^6.$$

We have  $\deg(f) = 3 = \deg(g)$ ,  $\deg(fg) = 6 = \deg(f) + \deg(g)$ , and  $\deg(f + g) = 2 \leq \max(\deg(f), \deg(g))$ . We see that strict inequality occurs in the last formula, since the leading coefficients of  $f$  and  $g$  cancel in  $f + g$ .

We stress once more that formal polynomials are not “functions of  $x$ .” Two formal polynomials  $f = \sum_{n \geq 0} f_n x^n$  and  $g = \sum_{n \geq 0} g_n x^n$  are equal iff  $f_n = g_n$  for all  $n \in \mathbb{N}$ . This is true by the definition of formal polynomials as sequences (functions with domain  $\mathbb{N}$ ) and is equivalent to the linear independence of the basis  $B$ . Nevertheless, we can use a formal polynomial to define an associated polynomial function, as follows.

**7.17. Definition: Polynomial Functions.** Given a nonzero polynomial  $f \in K[x]$  and a commutative ring  $R$  containing the field  $K$ , the *polynomial function associated to  $f$  with domain  $R$*  is the function  $P_f : R \rightarrow R$  defined by

$$P_f(z) = \sum_{n=0}^{\deg(f)} f_n z^n \quad (z \in R).$$

If  $f = 0$ , we let  $P_f(z) = 0$  for all  $z \in R$ .

One can show that, *because  $R$  contains the infinite field  $\mathbb{Q}$* ,  $f = g$  (equality of formal polynomials) iff  $P_f = P_g$  (equality of functions). However, this statement fails if one considers polynomial functions defined on *finite* rings and fields (see the exercises).

**7.18. Example.** Let  $f = 2 + x^2 + 3x^3 \in \mathbb{Q}[x]$ , and let  $R = \mathbb{C}$  (the complex numbers). Then

$$P_f(\sqrt{2}) = 2 + (\sqrt{2})^2 + 3(\sqrt{2})^3 = 4 + 6\sqrt{2}; \quad P_f(i) = 2 + i^2 + 3i^3 = 1 - 3i.$$

**7.19. Example.** Let  $h = 1 - x + x^2 \in \mathbb{Q}[x]$ , and let  $R = \mathbb{Q}[x]$ . Then  $P_h(2x^3) = 1 - (2x^3) + (2x^3)^2 = 1 - 2x^3 + 4x^6$  and  $P_h(h) = 1 - (1 - x + x^2) + (1 - x + x^2)^2 = 1 - x + 2x^2 - 2x^3 + x^4$ . Note also that  $P_h(x) = 1 - x + x^2 = h$ ; more generally,  $P_f(x) = f$  for any  $f \in K[x]$ . Next suppose  $R = \mathbb{Q}[[x]]$  and  $z = \sum_{n \geq 0} x^n \in R$ . Then

$$P_h(z) = (1, 0, 0, 0, 0, \dots) - (1, 1, 1, 1, 1, \dots) + (1, 2, 3, 4, 5, \dots) = (1, 1, 2, 3, 4, \dots).$$

Intuitively, the next result confirms that the algebraic operations on formal polynomials agree with the familiar algebraic operations on polynomial functions.

**7.20. Theorem: Comparison of Algebraic Operations on Formal Polynomials and Polynomial Functions.** Let  $f, g \in K[x]$ , let  $c \in K \subseteq K[x]$ , let  $R$  be a commutative ring containing the field  $K$ , and let  $z \in R$ . (a)  $P_{f+g}(z) = P_f(z) + P_g(z)$ . (b)  $P_{fg}(z) = P_f(z)P_g(z)$ . (c)  $P_c(z) = c$ .

*Proof.* We prove (b), leaving (a) and (c) as exercises. Both sides in (b) are zero if  $f = 0$  or  $g = 0$ , so assume  $f \neq 0$  and  $g \neq 0$ . Write  $n = \deg(f)$  and  $m = \deg(g)$ , so  $\deg(fg) = n + m$ .

Now, compute

$$\begin{aligned}
 P_{fg}(z) &= \sum_{k=0}^{n+m} (fg)_k z^k \text{ (definition of } P_{fg}) \\
 &= \sum_{k=0}^{n+m} \left( \sum_{i+j=k} f_i g_j \right) z^k \text{ (definition of } fg) \\
 &= \sum_{k=0}^{n+m} \sum_{i+j=k} (f_i z^i g_j z^j) \text{ (distributive law in } R \text{ and commutativity of } R) \\
 &= \left( \sum_{i=0}^n f_i z^i \right) \left( \sum_{j=0}^m g_j z^j \right) \text{ (generalized distributive law in } R) \\
 &= P_f(z) P_g(z) \text{ (definition of } P_f \text{ and } P_g). \quad \square
 \end{aligned}$$

We can rephrase this result in a somewhat more sophisticated way, using the following definition.

**7.21. Definition: Ring Homomorphisms.** Let  $R$  and  $S$  be rings. A map  $f : R \rightarrow S$  is a *ring homomorphism* iff  $f(1_R) = 1_S$  and for all  $x, y \in R$ ,  $f(x + y) = f(x) + f(y)$  and  $f(xy) = f(x)f(y)$ .

**7.22. Theorem: Evaluation Homomorphisms on  $K[x]$ .** Suppose  $R$  is a commutative ring containing the field  $K$ . For each  $z \in R$ , there exists a unique ring homomorphism  $\text{ev}_z : K[x] \rightarrow R$  such that  $\text{ev}_z(x) = z$  and  $\text{ev}_z(c) = c$  for all  $c \in K$ . Furthermore,  $\text{ev}_z(f) = P_f(z)$  for all  $f \in K[x]$ . We call  $\text{ev}_z$  the *evaluation homomorphism on  $K[x]$  determined by evaluating  $x$  at  $z$* .

*Proof.* The particular map  $\text{ev}_z(f) = P_f(z)$  is a ring homomorphism sending  $x$  to  $z$  and fixing  $K$ , since (by the previous theorem) for all  $f, g \in K[x]$  and  $c \in K$ ,

$$\begin{aligned}
 \text{ev}_z(f + g) &= P_{f+g}(z) = P_f(z) + P_g(z) = \text{ev}_z(f) + \text{ev}_z(g); \\
 \text{ev}_z(fg) &= P_{fg}(z) = P_f(z)P_g(z) = \text{ev}_z(f)\text{ev}_z(g); \\
 \text{ev}_z(c) &= P_c(z) = c \quad (\text{so } \text{ev}_z(1_{K[x]}) = 1_R);
 \end{aligned}$$

and  $\text{ev}_z(x) = P_x(z) = z$ . To prove uniqueness, let  $E : K[x] \rightarrow R$  be any ring homomorphism with  $E(x) = z$  and  $E(c) = c$  for  $c \in K$ . For a nonzero  $f \in K[x]$  of degree  $n$ , we have

$$E(f) = E\left(\sum_{k=0}^n f_k x^k\right) = \sum_{k=0}^n E(f_k x^k) = \sum_{k=0}^n E(f_k)E(x)^k = \sum_{k=0}^n f_k z^k = P_f(z) = \text{ev}_z(f).$$

As  $E(0) = 0_R = \text{ev}_z(0)$ , we conclude that  $E = \text{ev}_z$ .  $\square$

## 7.4 Order of Formal Power Series

We now discuss the order of a formal power series, which is analogous to the degree of a formal polynomial. The degree of a polynomial  $F$  is the largest  $n$  with  $F(n) \neq 0$ . Such an  $n$  will not exist if  $F$  is a formal power series that is not a polynomial. So we instead proceed as follows.

**7.23. Definition: Order of a Formal Power Series.** Let  $F = \sum_{n \geq 0} F_n x^n \in K[[x]]$  be a nonzero series. The *order* of  $F$ , denoted  $\text{ord}(F)$ , is the least  $n \geq 0$  such that  $F(n) \neq 0$ . Since  $\{k \in \mathbb{N} : F(k) \neq 0\}$  is a nonempty subset of  $\mathbb{N}$ , the order of  $F$  is well defined. The order of the zero series is undefined.

**7.24. Example.** The order of  $(0, 0, 0, 4, 2, 1, 4, 2, 1, 4, 2, 1, \dots)$  is 3. Nonzero polynomials have both a degree and an order; for instance,  $x^2 + x^5 + 3x^7$  has degree 7 and order 2.

The properties of order are analogous to the properties of degree.

**7.25. Theorem: Properties of Order.** Let  $F$  and  $G$  be nonzero formal series in  $K[[x]]$ .  
 (a) If  $F + G \neq 0$ , then  $\text{ord}(F + G) \geq \min(\text{ord}(F), \text{ord}(G))$ . (b)  $FG \neq 0$ , and  $\text{ord}(FG) = \text{ord}(F) + \text{ord}(G)$ .

*Proof.* Let  $n = \text{ord}(F)$  and  $m = \text{ord}(G)$ . (a) Let  $k = \min(n, m)$ . For any  $i < k$ ,  $(F + G)_i = F_i + G_i = 0 + 0 = 0$ , so  $\text{ord}(F + G) \geq k$ .

(b) For any  $p < n + m$ ,  $(FG)_p = \sum_{i+j=p} F_i G_j$ . For any pair  $(i, j) \in \mathbb{N}^2$  with sum  $p$ , either  $i < n$  or  $j < m$ . Thus  $F_i = 0$  or  $G_j = 0$ , so every summand  $F_i G_j = 0$ . Hence,  $(FG)_p = 0$ . On the other hand, for  $p = n + m$ , we only get a nonzero summand when  $i = n$  and  $j = m$ . So  $(FG)_{n+m} = F_n G_m \neq 0$ , since  $F_n \neq 0$  and  $G_m \neq 0$  and  $K$  is a field (cf. 7.27 below). This shows that  $FG \neq 0$  and that  $n + m$  is the least element  $p$  of  $\mathbb{N}$  with  $(FG)_p \neq 0$ , hence  $\text{ord}(FG) = n + m = \text{ord}(F) + \text{ord}(G)$ .  $\square$

**7.26. Definition: Integral Domains.** A commutative ring  $R$  with more than one element is an *integral domain* iff for all nonzero  $x, y \in R$ ,  $xy \neq 0$ . Equivalently, for all  $x, y \in R$ ,  $xy = 0$  implies  $x = 0$  or  $y = 0$ .

**7.27. Example.** Every field  $F$  (see 2.3) is an integral domain. For if  $x, y \in K$ ,  $xy = 0$ , and  $x \neq 0$ , then  $x$  has a multiplicative inverse in  $K$ . Multiplying  $xy = 0$  by the inverse of  $x$ , we see that  $y = 0$ . The ring  $\mathbb{Z}$  is an integral domain that is not a field. Part (b) of 7.14 shows that  $K[x]$  is an integral domain. Part (b) of 7.25 shows that  $K[[x]]$  is an integral domain. A key step in both proofs was the deduction that  $F_n G_m \neq 0$  since  $F_n \neq 0$  and  $G_m \neq 0$ . Thus, these proofs tacitly used the fact that the field  $K$  is an integral domain. This hypothesis on  $K$  (which is weaker than assuming that  $K$  is a field) is enough to ensure that  $K[x]$  and  $K[[x]]$  will be integral domains.

## 7.5 Formal Limits, Infinite Sums, and Infinite Products

Even in the algebraic setting of formal power series, one can imitate the limiting operations that play such a prominent role in calculus. In particular, we can use formal limits of formal power series to define infinite sums and infinite products of formal power series in certain situations.

**7.28. Definition: Limit of a Sequence of Formal Power Series.** Suppose  $(F_k : k \in \mathbb{N})$  is a sequence of elements of  $K[[x]]$  (so  $F_k : \mathbb{N} \rightarrow K$  for each  $k \in \mathbb{N}$ ), and  $G \in K[[x]]$ . Write

$$\lim_{k \rightarrow \infty} F_k = G \quad (\text{or } F_k \rightarrow G)$$

iff for each  $n \geq 0$ , there exists an index  $K(n)$  such that  $k \geq K(n)$  implies  $F_k(n) = G(n)$ .

Informally, the sequence  $(F_k : k \in \mathbb{N})$  of formal power series converges to some (necessarily unique) limit series in  $K[[x]]$  iff for each  $n \in \mathbb{N}$ , the coefficient of  $x^n$  in  $F_k$  eventually becomes constant for large enough values of  $k$  (and this constant is the coefficient of  $x^n$  in the limit series  $G$ ).

**7.29. Example.** We have  $\lim_{k \rightarrow \infty} x^k = 0$  in  $K[[x]]$ . To prove this, fix any  $n$  and then note that  $k > n$  implies  $x^k(n) = 0 = 0(n)$ .

Now that limits are available, we can define infinite sums (resp. products) as limits of partial finite sums (resp. products).

**7.30. Definition: Infinite Sums and Products of Formal Series.** Suppose  $(F_k : k \in \mathbb{N})$  is a sequence of formal power series in  $K[[x]]$ . For each  $N \geq 0$ , let  $G_N = F_0 + F_1 + \cdots + F_N \in K[[x]]$  be the  $N$ th partial sum of this sequence. If  $H = \lim_{N \rightarrow \infty} G_N$  exists in  $K[[x]]$ , then we write  $\sum_{k=0}^{\infty} F_k = H$ . Similarly, let  $P_N = F_0 F_1 \cdots F_N \in K[[x]]$  be the  $N$ th partial product of the sequence of  $F_k$ 's. If  $Q = \lim_{N \rightarrow \infty} P_N$  exists in  $K[[x]]$ , then we write  $\prod_{k=0}^{\infty} F_k = Q$ . Analogous definitions are made for sums and products ranging over any countably infinite index set (e.g., for  $k$  ranging from 1 to  $\infty$ ).

**7.31. Example.** Given  $F = \sum_{n=0}^{\infty} F_n x^n \in K[[x]]$ , define a formal series  $G_k = F_k X_k = F_k x^k$  (see 7.9) for each  $k \geq 0$ . We have

$$\sum_{k=0}^n G_k = \sum_{k=0}^n F_k x^k = (F_0, F_1, \dots, F_n, 0, 0, \dots).$$

Given  $m \in \mathbb{N}$ , it follows that the coefficient of  $x^m$  in any partial sum  $\sum_{k=0}^n G_k$  with  $n \geq m$  is  $F_m = F(m)$ . Thus, by definition,  $\sum_{k=0}^{\infty} G_k$  has limit  $F$ . In other words,

$$\sum_{k=0}^{\infty} F_k X_k = F = \sum_{k=0}^{\infty} F_k x^k,$$

where the left side is an infinite sum of formal power series, and the right side is our notation for the single formal power series  $F$ . This equality finally justifies the use of the “power series notation” for elements of  $K[[x]]$ .

The previous example can be rephrased as follows.

**7.32. Theorem: Density of Polynomials in  $K[[x]]$ .** For each  $F \in K[[x]]$ , there exists a sequence of polynomials  $f_n \in K[x]$  such that  $\lim_{n \rightarrow \infty} f_n = F$ . Specifically, we can take  $f_n = \sum_{k=0}^n F_k x^k$ .

Testing the convergence of infinite sums and products of real-valued functions is a delicate and often difficult problem. On the other hand, we can use the notion of order to give simple and convenient criteria ensuring the existence of infinite sums and infinite products of *formal* power series. Recall from calculus that a sequence of integers  $(e_n : n \geq 0)$  tends to infinity (in  $\mathbb{R}$ ) iff for every integer  $K > 0$ , there exists  $N$  such that  $n \geq N$  implies  $e_n > K$ .

**7.33. Theorem: Existence Criteria for Limits of Formal Series.** Suppose  $(F_k : k \in \mathbb{N})$  is a sequence of nonzero formal power series in  $K[[x]]$ .

- (a)  $\lim_{k \rightarrow \infty} F_k = 0$  in  $K[[x]]$  iff  $\lim_{k \rightarrow \infty} \text{ord}(F_k) = \infty$  in  $\mathbb{R}$ .
- (b)  $\sum_{k=0}^{\infty} F_k$  exists in  $K[[x]]$  iff  $\lim_{k \rightarrow \infty} \text{ord}(F_k) = \infty$  in  $\mathbb{R}$ .
- (c) If  $F_k(0) = 0$  for all  $k$ , then  $\prod_{k=0}^{\infty} (1 + F_k)$  exists in  $K[[x]]$  iff  $\lim_{k \rightarrow \infty} \text{ord}(F_k) = \infty$  in  $\mathbb{R}$ .

*Proof.* (a) Assume  $F_k \rightarrow 0$  in  $K[[x]]$ . Choose a fixed integer  $M \geq 0$ . For each  $n$  between 0 and  $M$ , there exists an index  $k_n$  such that  $k \geq k_n$  implies  $F_k(n) = 0$ . Hence, whenever  $k \geq K = \max(k_0, k_1, \dots, k_M)$ , we have  $F_k(n) = 0$  for all  $n \leq M$ . It follows that  $\text{ord}(F_k) > M$  whenever  $k \geq K$ . This proves that the sequence of integers  $(\text{ord}(F_k) : k \in \mathbb{N})$  tends to infinity as  $k$  goes to infinity. Conversely, suppose  $\text{ord}(F_k) \rightarrow \infty$  as  $k \rightarrow \infty$ . Fix  $n$ , and choose  $K$  so that  $k \geq K$  implies  $\text{ord}(F_k) > n$ . It follows immediately that  $F_k(n) = 0 = 0(n)$  for all  $k \geq K$ . Thus,  $F_k \rightarrow 0$  in  $K[[x]]$ .

(b) Suppose  $\sum_{k \geq 0} F_k$  converges to  $G$  in  $K[[x]]$ . Given an index  $n$ , we can therefore choose  $K$  so that  $k \geq K$  implies

$$(F_0 + F_1 + \dots + F_k)(n) = G(n).$$

Given  $k > K$ , note that

$$F_k(n) = (F_0 + \dots + F_k)(n) - (F_0 + \dots + F_{k-1})(n) = G(n) - G(n) = 0.$$

This proves that  $F_k \rightarrow 0$  as  $k \rightarrow \infty$ , and hence  $\text{ord}(F_k) \rightarrow \infty$  by (a). Conversely, suppose  $\text{ord}(F_k) \rightarrow \infty$ , so that  $F_k \rightarrow 0$  by (a). For each fixed  $n$ , the coefficient of  $x^n$  in  $F_k$  is eventually zero, and hence the coefficient of  $x^n$  in the partial sum  $F_0 + F_1 + \dots + F_k$  eventually stabilizes. Thus, these partial sums have a limit in  $K[[x]]$ .

(c) Suppose the indicated infinite product exists. We must show that for all  $n$ ,  $\text{ord}(F_k)$  is eventually  $\geq n$ . We prove this by induction on  $n$ . The statement is true for  $n = 1$  since  $F_k(0) = 0$  for all  $k$ . Assume the statement holds for some  $n \geq 1$ . Choose  $k_0$  so that  $\text{ord}(F_k) \geq n$  for all  $k \geq k_0$ . Next, using the hypothesis that the infinite product exists, choose  $k_1$  so that  $j, k \geq k_1$  implies

$$\left[ \prod_{i=0}^j (1 + F_i) \right] (n) = \left[ \prod_{i=0}^k (1 + F_i) \right] (n).$$

Note that

$$\prod_{i=0}^k (1 + F_i) - \prod_{i=0}^{k-1} (1 + F_i) = \left[ \prod_{i=0}^{k-1} (1 + F_i) \right] (1 + F_k - 1) = F_k \prod_{i=0}^{k-1} (1 + F_i). \quad (7.1)$$

For  $k > k_1$ , the coefficient of  $x^n$  on the left side is zero. On the other hand, for  $k \geq k_0$ , the fact that  $\text{ord}(F_k) \geq n$  implies that

$$\left[ F_k \prod_{i=0}^{k-1} (1 + F_i) \right] (n) = F_k(n),$$

since the partial product has constant term 1. Combining these facts, we see that  $F_k(n) = 0$  for  $k > \max(k_0, k_1)$ . This shows that  $\text{ord}(F_k)$  eventually exceeds  $n$ , completing the induction.

Conversely, suppose  $\text{ord}(F_k) \rightarrow \infty$  as  $k \rightarrow \infty$ . Fix  $n$ ; we must show that the coefficient of  $x^n$  in the partial products  $\prod_{i=0}^k (1 + F_i)$  eventually stabilizes. Choose  $k_0$  so that  $k \geq k_0$  implies  $\text{ord}(F_k) > n$ . It suffices to show that for all  $k > k_0$ ,

$$\left[ \prod_{i=0}^k (1 + F_i) \right] (n) = \left[ \prod_{i=0}^{k-1} (1 + F_i) \right] (n).$$

Subtracting and using (7.1), we see that the condition we want is equivalent to

$$\left[ F_k \prod_{i=0}^{k-1} (1 + F_i) \right] (n) = 0 \quad (k > k_0).$$

This holds because the product appearing on the left side here has order greater than  $n$ .  $\square$

**7.34. Example.** The infinite product  $\prod_{n=1}^{\infty} (1 + x^n)$  is a well-defined element of  $K[[x]]$ , since  $\text{ord}(x^n) = n \rightarrow \infty$  as  $n \rightarrow \infty$ .

**7.35. Theorem: Limit Rules for Sums and Products.** Suppose  $F_n, G_n, P, Q \in K[[x]]$  are formal series such that  $F_n \rightarrow P$  and  $G_n \rightarrow Q$ . Then  $F_n + G_n \rightarrow P + Q$  and  $F_n G_n \rightarrow PQ$ .

*Proof.* We prove the second statement, leaving the first as an exercise. Fix  $m \in \mathbb{N}$ . We must find  $N \in \mathbb{N}$  so that  $n \geq N$  implies  $(F_n G_n)(m) = (PQ)(m)$ . For each  $j \leq m$ , there is an  $N_j \in \mathbb{N}$  such that  $n \geq N_j$  implies  $F_n(j) = P(j)$ . Similarly, for each  $k \leq m$ , there is an  $M_k \in \mathbb{N}$  such that  $n \geq M_k$  implies  $G_n(k) = Q(k)$ . Let  $N = \max(N_0, \dots, N_m, M_0, \dots, M_m) \in \mathbb{N}$ . For any  $n \geq N$ ,

$$(PQ)(m) = \sum_{j+k=m} P(j)Q(k) = \sum_{j+k=m} F_n(j)G_n(k) = (F_n G_n)(m). \quad \square$$

## 7.6 Multiplicative Inverses in $K[x]$ and $K[[x]]$

In any ring  $S$ , it is of interest to know which elements of  $S$  have multiplicative inverses in  $S$ .

**7.36. Definition: Units of a Ring.** An element  $x$  in a ring  $S$  is called a *unit* of  $S$  iff there exists  $y \in S$  with  $xy = yx = 1_S$ .

Suppose  $y, z \in S$  satisfy  $xy = yx = 1_S$  and  $xz = zx = 1_S$ . Then  $y = y1 = y(xz) = (yx)z = 1z = z$ , so  $y = z$ . Thus, if  $x$  has a multiplicative inverse in  $S$ , this inverse is unique. We write  $x^{-1}$  or  $1/x$  to denote this inverse.

**7.37. Example.** If  $|S| > 1$ , then  $1_S \neq 0_S$ , and zero is not a unit of  $S$ . By definition, every nonzero element of a field  $F$  is a unit of  $F$ . In particular, the units of  $\mathbb{Q}$  are the nonzero rational numbers. On the other hand, the only units of the ring  $\mathbb{Z}$  are 1 and  $-1$ . So 2 is a unit of  $\mathbb{Q}$  but not a unit of  $\mathbb{Z}$ .

Next we characterize the units of the polynomial ring  $K[x]$  and the formal power series ring  $K[[x]]$ . Our first result says that the only units in the polynomial ring  $K[x]$  are the nonzero scalars.

**7.38. Theorem: Units of  $K[x]$ .** A polynomial  $f \in K[x]$  is a unit in  $K[x]$  iff  $\deg(f) = 0$ .

*Proof.* The zero polynomial is not a unit of  $K[x]$ , so assume  $f \neq 0$  henceforth. First, suppose  $f = a_0 x^0$  is a degree zero polynomial, so  $a_0 \in K$  is nonzero. Since  $K$  is a field,  $a_0^{-1}$  exists in  $K$ . In  $K[x]$ , we have  $a_0 a_0^{-1} = 1_{K[x]} = a_0^{-1} a_0$ , so  $a_0^{-1}$  is also a multiplicative inverse for  $f$  in the ring  $K[x]$ . Thus,  $f$  is a unit of  $K[x]$ .

Conversely, suppose  $\deg(f) > 0$ . For any nonzero  $g \in K[x]$ , we know  $\deg(fg) = \deg(f) + \deg(g) > 0$  (this result uses the fact that  $K$  is a field). Thus  $fg \neq 1_{K[x]}$ , since  $\deg(1_{K[x]}) = 0$ . So  $f$  does not have a multiplicative inverse  $g$  in  $K[x]$ .  $\square$

Intuitively, a non-constant polynomial cannot be a unit since there is no way to get rid of the positive powers of  $x$ . Perhaps surprisingly, when we pass to the larger ring of formal power series, almost every element in the ring becomes a unit. More precisely, every series with nonzero constant term has an inverse in  $K[[x]]$ . Before proving this, we consider an example that occurs frequently.

**7.39. Example: Formal Geometric Series.** Consider the series  $F = (1, -1, 0, 0, 0, \dots)$  and  $G = (1, 1, 1, 1, 1, \dots)$ . Multiplying these series, we discover that

$$FG = GF = (1, 0, 0, 0, \dots) = 1_{K[[x]]}.$$

Thus the polynomial  $1 - x$  is invertible in  $K[[x]]$ , and

$$(1 - x)^{-1} = 1 + x + x^2 + \dots + x^n + \dots = \sum_{n \geq 0} x^n.$$

This is a formal version of the “geometric series formula” learned in calculus. In calculus, one requires  $|x| < 1$  to ensure convergence. There is no such restriction here, since the letter  $x$  in our formula does not denote a real number!

**7.40. Theorem: Units in  $K[[x]]$ .** A formal power series  $F \in K[[x]]$  is a unit in  $K[[x]]$  iff  $F(0) \neq 0$ .

*Proof.* Assume  $F(0) = 0$ . For any  $G \in K[[x]]$ ,  $(FG)(0) = F(0)G(0) = 0 \neq 1 = 1_{K[[x]]}(0)$ . So  $FG \neq 1$  in  $K[[x]]$ , and  $F$  is not a unit of  $K[[x]]$ .

Conversely, assume  $F(0) \neq 0$ . Our goal is to find a series  $G = \sum_{n \geq 0} G_n x^n$  such that  $FG = GF = 1_{K[[x]]}$ . The desired equation  $FG = 1$  holds in  $K[[x]]$  iff the following infinite system of equations holds in the field  $K$ :

$$\begin{aligned} F_0 G_0 &= 1 \\ F_0 G_1 + F_1 G_0 &= 0 \\ F_0 G_2 + F_1 G_1 + F_2 G_0 &= 0 \\ \dots &\dots \\ \sum_{k=0}^n F_k G_{n-k} &= 0 \\ \dots &\dots \end{aligned} \tag{7.2}$$

We claim that there exist unique scalars  $G_0, G_1, \dots, G_n, \dots$  solving this system. To prove existence, we “solve the preceding system for the unknowns  $G_n$ .” More precisely, we recursively define  $G_0 = F_0^{-1} \in K$ ,  $G_1 = F_0^{-1}(-F_1 G_0)$ ,  $G_2 = F_0^{-1}(-F_1 G_1 - F_2 G_0)$ , and in general

$$G_n = -F_0^{-1} \sum_{k=1}^n F_k G_{n-k}. \tag{7.3}$$

By construction, the scalars  $G_n \in K$  defined in this way satisfy (7.2), and therefore  $G = \sum_{n \geq 0} G_n x^n$  satisfies  $FG = GF = 1$ . Since  $G = F^{-1}$  in  $K[[x]]$ ,  $G$  (and hence the  $G_n$ ) are uniquely determined by  $F$ .  $\square$

The preceding proof gives a recursive algorithm for calculating any given coefficient of  $1/F$  in terms of the coefficients of  $F$  and lower coefficients of  $1/F$ . In some situations, the following theorem (which generalizes 7.39) gives a more convenient way to calculate the multiplicative inverse of a formal series.

**7.41. Theorem: Formal Geometric Series.** If  $G \in K[[x]]$  satisfies  $G(0) = 0$ , then

$$(1 - G)^{-1} = \frac{1}{1 - G} = 1 + G + G^2 + G^3 + \dots = \sum_{k=0}^{\infty} G^k \in K[[x]].$$

*Proof.* The theorem holds if  $G = 0$ , so assume  $G$  is nonzero. Suppose  $\text{ord}(G) = d > 0$ ; then  $\text{ord}(G^k) = kd$ , which goes to infinity as  $k \rightarrow \infty$ . It follows that  $H = \sum_{k=0}^{\infty} G^k$  exists.

Consider the coefficient of  $x^n$  in  $(1 - G)H$ . By order considerations, this coefficient is the same as the coefficient of  $x^n$  in

$$(1 - G)(1 + G + G^2 + \cdots + G^n) = 1 - G^{n+1}.$$

Since  $G^{n+1}(n) = 0$ , we see that the coefficient in question is 1 if  $n = 0$  and is 0 if  $n > 0$ . Thus,  $(1 - G)H = H(1 - G) = 1$ , so  $H = (1 - G)^{-1}$ .  $\square$

**7.42. Example.** Taking  $G = x^i$  in the theorem (where  $i \geq 1$  is a fixed integer), we find that

$$\frac{1}{1 - x^i} = 1 + x^i + x^{2i} + x^{3i} + \cdots \in K[[x]].$$

This series has the form  $1 + F$  where  $\text{ord}(F) = i$ . It follows that the infinite product  $\prod_{i=1}^{\infty} \frac{1}{1 - x^i}$  exists. Similarly,  $\prod_{i=1}^{\infty} (1 - x^i)$  exists because  $\text{ord}(-x^i) = i$ , which goes to infinity as  $i$  goes to infinity. Next, we claim that

$$\prod_{i=1}^{\infty} (1 - x^i) \prod_{i=1}^{\infty} \frac{1}{1 - x^i} = 1.$$

One might at first believe that this identity is automatically true (due to “cancellation”), but care is required since we are dealing with formal infinite products. To justify this identity carefully, let  $P_n = \prod_{i=1}^n (1 - x^i)$  and  $Q_n = \prod_{i=1}^n (1 - x^i)^{-1}$  for each  $n \in \mathbb{N}$ . Using 7.35, we see that

$$\prod_{i=1}^{\infty} (1 - x^i) \prod_{i=1}^{\infty} \frac{1}{1 - x^i} = \left( \lim_{n \rightarrow \infty} P_n \right) \cdot \left( \lim_{n \rightarrow \infty} Q_n \right) = \lim_{n \rightarrow \infty} P_n Q_n = \lim_{n \rightarrow \infty} 1 = 1.$$

Note that we can rearrange the factors in each *finite* product  $P_n Q_n$  (since multiplication is commutative) and cancel to obtain 1.

**7.43. Example.** The geometric series formula can be used to invert any invertible formal power series, not just series with constant term 1. For, suppose  $F = \sum_{n \geq 0} F_n x^n$  where  $F_0 \neq 0$ . We can write  $F = F_0(1 - G)$  where  $G = \sum_{n \geq 1} (-F_0^{-1} F_n) x^n$ . Then

$$F^{-1} = F_0^{-1}(1 - G)^{-1} = F_0^{-1}[1 + G + G^2 + \cdots + G^k + \cdots].$$

## 7.7 Formal Laurent Series

We saw in the last section that  $G \in K[[x]]$  is a unit in  $K[[x]]$  iff  $G(0) \neq 0$ . Sometimes we want to divide by elements of  $K[[x]]$  that are not units. To do this, we need to operate in the *field of fractions* of the integral domain  $K[[x]]$ . We summarize this construction now, omitting many routine details; more thorough treatments may be found in texts on abstract algebra.

**7.44. Construction: Field of Fractions of an Integral Domain.** Let  $D$  be an integral domain, and let  $D^*$  be the set of nonzero elements of  $D$ . Let  $X = D \times D^*$  be the set of pairs  $(a, b)$  where  $a, b \in D$  and  $b$  is nonzero. Define a relation  $\sim$  on  $X$  by setting  $(a, b) \sim (c, d)$  iff  $ad = bc$ . One may verify that  $\sim$  is an equivalence relation (checking transitivity requires the assumption that  $D$  is an integral domain). Write  $F = \text{Frac}(D)$  to denote the set of



equivalence classes of this equivalence relation; also, write  $a/b$  to denote the equivalence class of  $(a, b)$ .

Given two elements  $a/b$  and  $c/d$  in  $F$ , define addition and multiplication as follows:

$$(a/b) + (c/d) = (ad + bc)/(bd); \quad (a/b) \times (c/d) = (ac)/(bd).$$

It must be checked that these operations are independent of the representatives chosen for the equivalence classes. For example, one must show that  $a/b = a'/b'$  and  $c/d = c'/d'$  imply  $(ad + bc)/(bd) = (a'd' + b'c')/(b'd')$ . Also define zero and one in  $F$  by  $0_F = 0_D/1_D$  and  $1_F = 1_D/1_D$ . One may now check that  $(F, +, \times, 0_F, 1_F)$  is a commutative ring with  $1_F \neq 0_F$ . The map  $i : D \rightarrow F$  such that  $i(a) = a/1_D$  for  $a \in D$  is an injective ring homomorphism that embeds  $D$  as a subring of  $F$  and allows us to regard  $D$  as a subset of  $F$ . Finally (and this is the point of the whole construction), every nonzero element  $a/b \in F$  has a multiplicative inverse in  $F$ , namely  $b/a$ . This follows since  $(a/b) \times (b/a) = (ab)/(ba)$ , and this equals  $1_D/1_D = 1_F$  because  $(ab)1 = (ba)1$  in  $D$ . Therefore,  $F$  is a field.

The field  $F$  has the following *universal mapping property*: for any ring homomorphism  $g : D \rightarrow L$  into a field  $L$ , there exists a unique ring homomorphism  $g' : F \rightarrow L$  extending  $g$  (more precisely, such that  $g = g' \circ i$ ). This homomorphism must be given by  $g'(a/b) = g(a)g(b)^{-1} \in L$  (proving uniqueness); for existence, one checks that the formula just written does give a well-defined ring homomorphism extending  $g$ .

**7.45. Example:  $\mathbb{Z}$  and  $\mathbb{Q}$ .** Let  $\mathbb{Z}$  be the ring of integers, which is an integral domain. The field  $\mathbb{Q}$  of rational numbers is, by definition, the field of fractions of  $\mathbb{Z}$ .

**7.46. Definition: Formal Rational Functions.** The symbol  $K(x)$  denotes the fraction field of the integral domain  $K[x]$ . Elements of  $K(x)$  are *formal rational functions*  $p/q$ , where  $p, q$  are polynomials with  $q$  nonzero; we have  $p/q = s/t$  in  $K(x)$  iff  $pt = qs$  in  $K[x]$ .

**7.47. Definition: Formal Laurent Series.** The symbol  $K((x))$  denotes the fraction field of the integral domain  $K[[x]]$ . Elements of  $K((x))$  are *formal Laurent series* in one indeterminate. A formal Laurent series is a quotient  $G/H$ , where  $G, H$  are formal power series with  $H$  nonzero; we have  $G/H = P/Q$  in  $K((x))$  iff  $GQ = HP$  in  $K[[x]]$ .

We can use our characterization of units in  $K[[x]]$  to find a canonical description of the elements of  $K((x))$ .

**7.48. Theorem: Representation of Laurent Series.** For every nonzero  $S \in K((x))$ , there exists a unique integer  $N$  and a unique series  $F \in K[[x]]$  such that  $S = x^N F$  and  $F(0) \neq 0$ .

**7.49. Remark.** We call  $N$  the *order* of  $S$ . When  $N \geq 0$ , so that  $S \in K[[x]]$ , this is consistent with the previous definition of  $\text{ord}(S)$ . When  $N = -m$  is negative,  $S = x^N F$  is the fraction  $F/x^m$ . In this case, we often use the “Laurent series notation”

$$S = F_0 x^{-m} + F_1 x^{-m+1} + F_2 x^{-m+2} + \cdots + F_m x^0 + F_{m+1} x^1 + \cdots = \sum_{n=-m}^{\infty} F_{n+m} x^n.$$

*Proof.* Given a nonzero  $S \in K((x))$ , there exist nonzero series  $G, H \in K[[x]]$  with  $S = G/H$ ;  $G$  and  $H$  are not unique. Write  $\text{ord}(G) = i$ ,  $\text{ord}(H) = j$ ,  $G = \sum_{n \geq i} G_n x^n$ ,  $H = \sum_{n \geq j} H_n x^n$ , where  $G_i$  and  $H_j$  are nonzero. Let  $H^* = \sum_{n \geq 0} H_{n+j} x^n$ , which is a unit in  $K[[x]]$  since  $H^*(0) = H_j \neq 0$ . Let  $Q \in K[[x]]$  be the inverse of  $H^*$ , which satisfies  $Q(0) \neq 0$ . Similarly, write  $G^* = \sum_{n \geq 0} G_{n+i} x^n$  and note that  $G^*(0) \neq 0$ . Now,

$$S = \frac{G}{H} = \frac{x^i G^*}{x^j H^*} = \frac{x^i G^* Q}{x^j H^* Q} = x^N F$$

if we set  $N = i - j \in \mathbb{Z}$  and  $F = G^*Q \in K[[x]]$ . This proves existence of  $N$  and  $F$ .

For uniqueness, assume  $x^N F = x^M P$  for some  $M \in \mathbb{Z}$  and some  $P \in K[[x]]$  with  $P(0) \neq 0$ . Choose  $k > \max(|N|, |M|)$ ; then  $x^{k+N} F = x^{k+M} P$ . Both sides are nonzero series in  $K[[x]]$  and hence have an order. Since  $F$  and  $P$  have nonzero constant term, comparison of orders gives  $k + N = k + M$  and hence  $N = M$ . Dividing by  $x^N$  (which is a unit in  $K((x))!$ ), we see that  $F = P$ .  $\square$

**7.50. Remark.** The proof shows that, for *any* representation of a nonzero  $S \in K((x))$  as a quotient  $F/G$  with  $F, G \in K[[x]]$ , we have  $\text{ord}(S) = \text{ord}(F) - \text{ord}(G)$ .

## 7.8 Formal Derivatives

We now define a formal version of the derivative operation studied in calculus.

**7.51. Definition: Formal Derivatives.** Given a formal series  $F = \sum_{n \geq 0} F_n x^n \in K[[x]]$ , the *formal derivative* of  $F$  is

$$F' = \frac{dF}{dx} = DF = \sum_{n \geq 0} (n+1) F_{n+1} x^n.$$

Higher-order formal derivatives are defined recursively by setting  $F^{(k+1)} = (F^{(k)})'$  for  $k \geq 1$ . It follows that

$$F^{(k)} = \sum_{n \geq 0} (n+1)(n+2) \cdots (n+k) F_{n+k} x^n.$$

The integer coefficients appearing in these formulas make sense, since we have assumed that  $K$  is a field containing  $\mathbb{Q}$ .

To give examples of formal differentiation, we introduce formal versions of some familiar functions from calculus.

**7.52. Definition: Formal Versions of Exponential, Logarithmic, and Trigonometric Functions.** Define the following elements in  $K[[x]]$  (recall  $K$  contains  $\mathbb{Q}$ ):

$$e^x = (1, 1, 1/2, 1/6, 1/24, 1/120, \dots) = \sum_{n \geq 0} \frac{1}{n!} x^n;$$

$$\sin x = (0, 1, 0, -1/6, 0, 1/120, 0, \dots) = \sum_{n \geq 0} \chi(n \text{ is odd}) \frac{(-1)^{(n-1)/2}}{n!} x^n;$$

$$\cos x = (1, 0, -1/2, 0, 1/24, 0, -1/720, \dots) = \sum_{n \geq 0} \chi(n \text{ is even}) \frac{(-1)^{n/2}}{n!} x^n;$$

$$\log(1+x) = (0, 1, -1/2, 1/3, -1/4, 1/5, \dots) = \sum_{n \geq 1} \frac{(-1)^{n+1}}{n} x^n.$$

**7.53. Example.** Let  $F = e^x$ ,  $G = \sin x$ ,  $H = \cos x$ , and  $P = \log(1+x)$  in  $K[[x]]$ . Using the definition of formal derivatives, we find that

$$F' = (1 \cdot 1, 2 \cdot (1/2), 3 \cdot (1/3!), \dots, (n+1) \cdot (1/(n+1)!), \dots) = (1, 1, 1/2, 1/6, \dots, 1/n!, \dots) = F.$$

Thus the formal power series  $e^x$  equals its own derivative, in accordance with the situation in calculus. Iterating, we see that  $f^{(k)} = f$  for all  $k \geq 1$ . Similar calculations show that  $G' = H$ ,  $H' = -G$ , and  $P' \cdot (1, 1, 0, 0, \dots) = 1_{K[[x]]}$ . We can express the last fact by writing

$$\frac{d}{dx} \log(1+x) = (1+x)^{-1}.$$

Formal derivatives obey many of the same differentiation rules that ordinary derivatives satisfy. However, each of these rules must be reproved in the formal setting. Some of these rules are stated in the next theorem.

**7.54. Theorem: Formal Differentiation Rules.** Let  $F, G, H_n \in K[[x]]$  and  $c, c_n \in K$ .

- (a)  $(F + G)' = F' + G'$  (sum rule).
- (b)  $(cF)' = c(F')$  (scalar rule).
- (c) For  $N \in \mathbb{N}$ ,  $\left(\sum_{n=1}^N c_n H_n\right)' = \sum_{n=1}^N c_n H'_n$  (linear combination rule).
- (d)  $\frac{d}{dx}(x^k) = kx^{k-1}$  for all  $k \geq 0$  (power rule).
- (e)  $(FG)' = F(G') + (F')G$  (product rule).
- (f) If  $H_n \rightarrow F$ , then  $H'_n \rightarrow F'$  (derivative of a limit).
- (g) If  $S = \sum_{n=1}^{\infty} H_n$  exists, then  $S' = \sum_{n=1}^{\infty} H'_n$  (derivative of an infinite sum).

*Proof.* We prove (d), (e), and (f), leaving the others as exercises.

- (d) Recall that  $x^k = \sum_{n \geq 0} \chi(n=k)x^n$ . The definition of formal derivative gives

$$\frac{d}{dx}(x^k) = \sum_{n \geq 0} (n+1)\chi(n+1=k)x^n = \sum_{n \geq 0} k\chi(n=k-1)x^n = kx^{k-1}.$$

- (e) Note on the one hand that

$$(FG)'_m = (m+1) \cdot (FG)_{m+1} = (m+1) \sum_{k=0}^{m+1} F_k G_{m+1-k}.$$

On the other hand,

$$\begin{aligned} (FG' + F'G)_m &= (FG')_m + (F'G)_m \\ &= \sum_{k=0}^m F_k G'_{m-k} + \sum_{j=0}^m F'_j G_{m-j} \\ &= \sum_{k=0}^m F_k (m+1-k) G_{m+1-k} + \sum_{j=0}^m (j+1) F_{j+1} G_{m-j}. \end{aligned}$$

In the first summation, we can let  $k$  go from 0 to  $m+1$  (which adds a zero term). In the second summation, change the summation variable to  $k = j+1$  and add a zero term corresponding to  $k=0$ . We get

$$\begin{aligned} (FG' + F'G)_m &= \sum_{k=0}^{m+1} (m+1-k) F_k G_{m+1-k} + \sum_{k=0}^{m+1} k F_k G_{m+1-k} \\ &= \sum_{k=0}^{m+1} (m+1) F_k G_{m+1-k} = (FG)'_m. \end{aligned}$$

This completes the proof of the product rule.

- (f) Assume  $\lim_{i \rightarrow \infty} H_i = F$ . To prove  $\lim_{i \rightarrow \infty} H'_i = F'$ , fix  $m \in \mathbb{N}$ . Choose  $N$  so that  $n \geq N$  implies  $H_n(m+1) = F(m+1)$ . By definition of formal derivatives,  $n \geq N$  implies

$$H'_n(m) = (m+1)H_n(m+1) = (m+1)F(m+1) = F'(m). \quad \square$$

A formal version of the chain rule will be given shortly, once we define formal composition of two formal power series. We turn next to a formal analogue of the Maclaurin series of a real-valued function.

**7.55. Theorem: Formal Maclaurin Series.** For all  $F \in K[[x]]$ ,  $F = \sum_{k \geq 0} \frac{F^{(k)}(0)}{k!} x^k$ .

*Proof.* We have  $F^{(k)} = \sum_{n \geq 0} (n+1)(n+2) \cdots (n+k) F_{n+k} x^n$ , so  $F^{(k)}(0) = k! F_k$ . Since  $K$  contains  $\mathbb{Q}$ , we can divide both sides by  $k!$  in  $K$ . Thus,  $F_k = F^{(k)}(0)/k!$  for all  $k \in \mathbb{N}$ .  $\square$

## 7.9 Composition of Polynomials

Given functions  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ , we can form the composite function  $g \circ f : \mathbb{R} \rightarrow \mathbb{R}$  by setting  $(g \circ f)(z) = g(f(z))$  for each  $z \in \mathbb{R}$ . We would like to introduce a version of this composition operation for formal power series. There is an immediate difficulty, since the formal series  $F = \sum_{n \geq 0} F_n x^n \in K[[x]]$  is not a function of  $x$  and need not correspond to a convergent power series in the variable  $x$ . On the other hand, we saw in 7.17 that a polynomial  $f \in K[x]$  can be viewed as a function  $P_f : R \rightarrow R$ , where  $R$  is any commutative ring containing  $K$ . So we can define the formal composition of two *polynomials* as follows.

**7.56. Definition: Composition of Polynomials.** Given  $f, g \in K[x]$ , the *formal composition* of  $f$  and  $g$  is  $f \bullet g = P_f(g)$ . More explicitly, if  $f = \sum_{k=0}^n f_k x^k$  and  $g = \sum_{j=0}^m g_j x^j$ , then

$$f \bullet g = \sum_{k=0}^n f_k \left( \sum_{j=0}^m g_j x^j \right)^k \in K[x].$$

Note that the filled circle  $\bullet$  denotes formal composition, whereas an open circle  $\circ$  denotes ordinary composition of functions. We can also write  $f \bullet g = \text{ev}_g(f)$ , where  $\text{ev}_g : K[x] \rightarrow K[x]$  is the evaluation homomorphism that sets  $x$  equal to  $g$  (see 7.22).

The following theorem shows that formal composition of polynomials satisfies properties analogous to those satisfied by ordinary composition of functions. The proof makes heavy use of the properties of evaluation homomorphisms.

**7.57. Theorem: Properties of Polynomial Composition.** Suppose  $f, g, h \in K[x]$  are polynomials,  $c \in K$ , and  $R$  is a commutative ring containing  $K$ .

- (a)  $P_{f \bullet g} = P_f \circ P_g : R \rightarrow R$  (comparison of formal and ordinary composition).
- (b)  $(f \bullet g) \bullet h = f \bullet (g \bullet h)$  (associativity).
- (c)  $(f + g) \bullet h = (f \bullet h) + (g \bullet h)$ ,  $(f \cdot g) \bullet h = (f \bullet h) \cdot (g \bullet h)$ ,  $c \bullet h = c$ , and  $(cf) \bullet h = c(f \bullet h)$  (homomorphism properties).
- (d)  $(f \bullet g)' = (f' \bullet g) \cdot g'$  (formal chain rule).

*Proof.* (a) Given  $z \in R$ , we must show that  $P_{f \bullet g}(z) = P_f(P_g(z))$ . Let us rewrite each side in terms of evaluation homomorphisms. The left side is

$$P_{f \bullet g}(z) = \text{ev}_z(f \bullet g) = \text{ev}_z(\text{ev}_g(f)) = (\text{ev}_z \circ \text{ev}_g)(f).$$

The right side is

$$P_f(P_g(z)) = P_f(\text{ev}_z(g)) = \text{ev}_{\text{ev}_z(g)}(f).$$

The equality in (a) will follow if we can show that the two ring homomorphisms  $\text{ev}_z \circ \text{ev}_g$

and  $\text{ev}_{\text{ev}_z(g)}$  from  $K[x]$  into  $R$  are equal. By the uniqueness part of 7.22, we need only verify that the two homomorphisms have the same effect on the polynomial  $x$ . This holds because

$$(\text{ev}_z \circ \text{ev}_g)(x) = \text{ev}_z(\text{ev}_g(x)) = \text{ev}_z(g) = \text{ev}_{\text{ev}_z(g)}(x).$$

(b) In part (a), let  $R$  be the ring  $K[x]$  and apply both sides to  $z = h$ . We obtain

$$(f \bullet g) \bullet h = P_{f \bullet g}(h) = P_f(P_g(h)) = P_f(g \bullet h) = f \bullet (g \bullet h).$$

(c) This is a transcription of the statement that  $\text{ev}_h$  is a ring homomorphism fixing all elements of  $K$ ; for example,

$$(fg) \bullet h = \text{ev}_h(fg) = \text{ev}_h(f) \text{ev}_h(g) = (f \bullet h)(g \bullet h).$$

(d) Let us first prove the special case where  $f = x^n$  for some  $n \geq 0$ . Since  $f' = nx^{n-1}$ , we must show that

$$(g^n)' = ng^{n-1}g' \quad (g \in K[x]).$$

We proceed by induction on  $n \geq 0$ . The base case  $n = 0$  holds because both sides are zero. Assuming the formula holds for some  $n \geq 0$ , we use the formal product rule to calculate

$$(g^{n+1})' = (g^n g)' = (g^n)'g + g^n g' = ng^{n-1}g'g + g^n g' = (n+1)g^n g'.$$

The general case of (d) now follows because the formula is “ $K$ -linear in  $f$ .” More precisely, if (d) holds for polynomials  $f_1$  and  $f_2$ , it also holds for  $f_1 + f_2$  because (by (c))

$$\begin{aligned} ((f_1 + f_2) \bullet g)' &= ((f_1 \bullet g) + (f_2 \bullet g))' = (f_1 \bullet g)' + (f_2 \bullet g)' \\ &= (f_1' \bullet g)g' + (f_2' \bullet g)g' = (f_1' \bullet g + f_2' \bullet g)g' \\ &= ((f_1' + f_2') \bullet g)g' = ((f_1 + f_2)' \bullet g)g'. \end{aligned}$$

Similarly, if (d) holds for some  $f \in K[x]$ , then (d) holds for  $cf$  whenever  $c \in K$ . Since every polynomial is a finite  $K$ -linear combination of powers of  $x$ , it follows that (d) is true for all polynomials  $f$ , as desired.  $\square$

## 7.10 Composition of Formal Power Series

We can extend the definition of the formal composition  $f \bullet g$  to the case where  $f \in K[x]$  is a polynomial and  $G \in K[[x]]$  is a formal series by setting  $f \bullet G = P_f(G) = \text{ev}_G(f)$ , as in 7.56. A more challenging problem is to define the composition  $F \bullet G$  when  $F$  and  $G$  are both formal series. To see what can go wrong, suppose  $F = \sum_{n \geq 0} F_n x^n$  and  $G = \sum_{m \geq 0} G_m x^m$  are formal series. By analogy with the preceding definitions, we would like to define

$$F \bullet G = \sum_{n \geq 0} F_n G^n = \sum_{n \geq 0} F_n \left( \sum_{k \geq 0} G_k x^k \right)^n \in K[[x]].$$

The trouble is that the infinite sum of formal series  $\sum_{n \geq 0} F_n G^n$  may not be well-defined. Indeed, if  $F_n \neq 0$  for infinitely many values of  $n$  and  $G_0 = 1$ , consideration of the constant term shows that  $\sum_{n \geq 0} F_n G^n$  does not exist. However, we can escape this difficulty by requiring that the right-hand factor in a formal composition  $F \bullet G$  have zero constant term. This leads to the following definition.

**7.58. Definition: Composition of Formal Power Series.** Given  $F, G \in K[[x]]$  with  $G(0) = 0$ , the *formal composition of  $F$  and  $G$*  is

$$F \bullet G = \sum_{n \geq 0} F_n G^n \in K[[x]].$$

This infinite sum converges by 7.33, because  $\text{ord}(F_n G^n) \geq n$  whenever  $F_n G^n \neq 0$  and so the orders of the nonzero summands go to  $\infty$  as  $n \rightarrow \infty$ .

**7.59. Theorem: Identity for Formal Composition.** For all  $F \in K[[x]]$ ,  $F \bullet x = F$ . For all  $G \in K[[x]]$  with  $G(0) = 0$ ,  $x \bullet G = G$ .

*Proof.* For any  $F \in K[[x]]$ ,  $F \bullet x = \sum_{n \geq 0} F_n x^n = F$ . Next, recall that  $x = X_1 = \sum_{n \geq 0} \chi(n=1)x^n$ . So, for  $G \in K[[x]]$  with  $G(0) = 0$ ,

$$x \bullet G = \sum_{n \geq 0} \chi(n=1)G^n.$$

One may check that this infinite sum of formal series converges to  $G^1 = G$ .  $\square$

The next technical result will aid us in proving further facts about formal composition.

**7.60. Theorem: Coefficient in a Composition.** For all  $F, G \in K[[x]]$  with  $G(0) = 0$  and all  $m \in \mathbb{N}$ ,

$$(F \bullet G)_m = \left( \sum_{n=0}^m F_n G^n \right)_m.$$

*Proof.* Since  $(F \bullet G)_m = (\sum_{n=0}^{\infty} F_n G^n)_m$ , it suffices to show that

$$\left( \sum_{n=0}^p F_n G^n \right)_m = \left( \sum_{n=0}^{p+1} F_n G^n \right)_m$$

for all  $p \geq m$ . This holds since

$$\left( \sum_{n=0}^{p+1} F_n G^n \right)_m = \left( \sum_{n=0}^p F_n G^n \right)_m + (F_{p+1} G^{p+1})_m,$$

and  $F_{p+1} G^{p+1}$  is either zero or has order at least  $p+1 > m$  (since  $\text{ord}(G) \geq 1$ ).  $\square$

**7.61. Theorem: Joint Continuity of Formal Composition.** Suppose  $F_n, G_n, P, Q \in K[[x]]$  are formal series such that  $G_n(0) = 0$  for all  $n \in \mathbb{N}$ ,  $F_n \rightarrow P$ , and  $G_n \rightarrow Q$  (forcing  $Q(0) = 0$ ). Then  $F_n \bullet G_n \rightarrow P \bullet Q$ .

*Proof.* Fix  $m \in \mathbb{N}$ ; we must show that  $(P \bullet Q)(m) = (F_n \bullet G_n)(m)$  for all sufficiently large  $n$ . By 7.60,

$$(P \bullet Q)(m) = \left( \sum_{i=0}^m P(i) Q^i \right)_m; \quad (7.4)$$

$$(F_n \bullet G_n)(m) = \left( \sum_{i=0}^m F_n(i) G_n^i \right)_m. \quad (7.5)$$

Now, for each fixed  $i \leq m$ , iteration of 7.35 shows that  $G_n^i \rightarrow Q^i$  as  $n \rightarrow \infty$ . For each

fixed  $i$ , the sequence of (order zero) power series  $F_n(i)$  converges to  $P(i)$  as  $n \rightarrow \infty$ , since  $F_n \rightarrow P$ . Using 7.35 again, we see that  $F_n(i)G_n^i \rightarrow P(i)Q^i$  for each fixed  $i \leq m$ , and hence (by 7.35)

$$\lim_{n \rightarrow \infty} \left( \sum_{i=0}^m F_n(i)G_n^i \right) = \sum_{i=0}^m P(i)Q^i.$$

It follows that the right sides of (7.4) and (7.5) do agree for large enough  $n$ , which is what we needed to show.  $\square$

The previous result combined with 7.32 allows us to use “continuity arguments” to deduce properties of composition of formal power series from corresponding properties of composition of polynomials. The next few theorems illustrate this technique.

**7.62. Theorem: Homomorphism Properties of Formal Composition.**

Let  $G \in K[[x]]$  satisfy  $G(0) = 0$ .

(a) For all  $F, H \in K[[x]]$ ,  $(F + H) \bullet G = (F \bullet G) + (H \bullet G)$ .

(b) For all  $F, H \in K[[x]]$ ,  $(FH) \bullet G = (F \bullet G)(H \bullet G)$ .

(c) For all  $c \in K$ ,  $c \bullet G = c$ .

So, the *evaluation map*  $\text{ev}_G : K[[x]] \rightarrow K[[x]]$  given by  $\text{ev}_G(F) = F \bullet G$  is a ring homomorphism fixing  $K$  and sending  $x$  to  $G$ .

*Proof.* We prove (a), leaving (b) and (c) as exercises. Fix  $F, G, H \in K[[x]]$  with  $G(0) = 0$ . Use 7.32 to choose polynomials  $f_n, g_n, h_n \in K[x]$  with  $g_n(0) = 0$  for all  $n$ ,  $f_n \rightarrow F$ ,  $g_n \rightarrow G$ , and  $h_n \rightarrow H$ . For each  $n \in \mathbb{N}$ , we know from 7.57(c) that

$$(f_n + h_n) \bullet g_n = (f_n \bullet g_n) + (h_n \bullet g_n).$$

Take the limit of both sides as  $n \rightarrow \infty$ . Using 7.35 and 7.61, we get  $(F + H) \bullet G = (F \bullet G) + (H \bullet G)$  as desired.  $\square$

**7.63. Theorem: Associativity of Formal Composition.** Suppose  $F, G, H \in K[[x]]$  satisfy  $G(0) = 0 = H(0)$ . Then

$$(F \bullet G) \bullet H = F \bullet (G \bullet H).$$

*Proof.* First note that all compositions in the theorem statement are defined; in particular,  $F \bullet (G \bullet H)$  is defined because  $G \bullet H$  has zero constant term. Use 7.32 to choose polynomials  $f_n, g_n, h_n \in K[x]$  with  $g_n(0) = 0 = h_n(0)$  for all  $n$ ,  $f_n \rightarrow F$ ,  $g_n \rightarrow G$ , and  $h_n \rightarrow H$ . For each  $n \in \mathbb{N}$ , we know from 7.57(b) that  $(f_n \bullet g_n) \bullet h_n = f_n \bullet (g_n \bullet h_n)$ . Taking limits and using 7.61 repeatedly gives the desired result.  $\square$

A similar continuity argument (left as an exercise) establishes the following differentiation rule.

**7.64. Theorem: Formal Chain Rule.** For all  $F, G \in K[[x]]$  with  $G(0) = 0$ ,

$$(F \bullet G)' = (F' \bullet G)G'.$$

**7.65. Theorem: Inverses for Formal Composition.** Let  $S = \{F \in K[[x]] : F(0) = 0 \text{ and } F(1) \neq 0\}$ .

(a) If  $F, G \in S$ , then  $F \bullet G \in S$  (closure of  $S$ ).

(b) If  $F \in S$ , there exists a unique  $G \in S$  with  $F \bullet G = x = G \bullet F$  (inverses).

(Together with 7.59 and 7.63, this proves that  $(S, \bullet)$  is a *group* as defined in 9.1. The proof will show that if  $F, G \in S$  and  $G \bullet F = x$ , then  $F \bullet G = x$  automatically follows.)

*Proof.* (a) Suppose  $F$  and  $G$  belong to  $S$ . On one hand, since  $F(0) = 0 = G(0)$ ,  $F \bullet G$  is defined and also has zero constant term. On the other hand, 7.60 gives  $(F \bullet G)_1 = F_1 G_1^1 \neq 0$ . So  $F \circ G \in S$ .

(b) First we prove that for each  $F \in S$ , there exists a unique  $G \in S$  with  $G \bullet F = x$  (we call  $G$  a “left inverse” of  $F$ ). By 7.60,

$$(G \bullet F)_n = \left( \sum_{m=0}^n G_m F^m \right)_n \quad (n \in \mathbb{N}).$$

We can use this equation to give a recursive prescription for the coefficients  $G_n$ . At the start, we must set  $G_0 = 0$  and  $G_1 = 1/F_1 \neq 0$ . (Note  $1/F_1$  exists because  $F_1$  is a nonzero element of the field  $K$ .) Assume  $n > 1$  and  $G_0, G_1, \dots, G_{n-1}$  have already been determined. Since  $(G_n F^n)_n = G_n (F_1)^n$ , we need to choose  $G_n$  so that

$$0 = \left( \sum_{m=0}^n G_m F^m \right)_n = G_n F_1^n + \left( \sum_{m=0}^{n-1} G_m F^m \right)_n.$$

Evidently there is a unique  $G_n \in K$  that will work, namely

$$G_n = -\frac{1}{F_1^n} \left( \sum_{m=0}^{n-1} G_m F^m \right)_n. \quad (7.6)$$

Since  $G \in S$ , we have shown that  $F$  has a unique left inverse in  $S$ .

To finish the proof, fix  $F \in S$ . Let  $G$  be the left inverse of  $F$ , and let  $H$  be the left inverse of  $G$ . Then

$$H = H \bullet x = H \bullet (G \bullet F) = (H \bullet G) \bullet F = x \bullet F = F.$$

Since  $H = F$ , we see that both  $G \bullet F$  and  $F \bullet G = H \bullet G$  equal the identity element  $x$ . Thus,  $G$  is the two-sided inverse of  $F$ .  $\square$

**7.66. Remark.** Lagrange’s inversion formula (8.15) provides an alternate way to determine the coefficients of the compositional inverse of a formal series  $F$ , which is sometimes easier to use than the recursive formula for  $G_n$  in the preceding proof.

**7.67. Example.** Consider the series  $E = e^x - 1 = \sum_{n \geq 1} x^n/n!$  and  $L = \log(1 + x) = \sum_{n \geq 1} (-1)^{n-1} x^n/n$ . Let us show that  $L$  is the two-sided inverse of  $E$  relative to formal composition. Set  $H = L \bullet E$ ; since  $H(0) = 0$ , it will suffice to prove that  $H' = 1$  (cf. 7.132). First, a routine formal differentiation shows that  $E' = e^x = 1 + E$  and  $L' = 1 - x + x^2 - x^3 + \dots = (1 + x)^{-1}$ . We also have  $L' \bullet E = 1 - E + E^2 - E^3 + \dots = (1 + E)^{-1}$  by 7.41. The formal chain rule now gives

$$H' = (L \bullet E)' = (L' \bullet E)E' = (1 + E)^{-1}(1 + E) = 1.$$

We conclude that  $L \bullet E = x$ , hence also  $E \bullet L = x$ .

## 7.11 Generalized Binomial Expansion

Given a formal power series  $F \in K[[x]]$ , we would like to give meaning to expressions like  $F^{1/2} = \sqrt{F}$ . To prepare for this, we will first define, for each  $r \in K$ , a power series  $\text{Pow}_r \in K[[x]]$  that is a formal analogue of the function  $x \mapsto (1+x)^r$ . The following example from calculus motivates the definition of  $\text{Pow}_r$ .



**7.68. Example.** Consider the real-valued function  $f(x) = (1+x)^r$ , where  $r$  is a fixed real constant, and  $x$  ranges over real numbers  $> -1$ . If  $f$  has a Taylor series expansion about  $x = 0$ , then the coefficients of the power series must be given by Taylor's formula

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$$

(cf. 7.55). Computing the successive derivatives of  $f$ , we find  $f'(x) = r(1+x)^{r-1}$ ,  $f''(x) = r(r-1)(1+x)^{r-2}$ , and in general

$$f^{(n)}(x) = r(r-1)(r-2) \cdots (r-n+1)(1+x)^{r-n} = (r)_{\downarrow n} (1+x)^{r-n}$$

(here we use the falling factorial notation from 2.76). Evaluating the derivatives at  $x = 0$ , we conclude that

$$(1+x)^r = \sum_{n=0}^{\infty} \frac{(r)_{\downarrow n}}{n!} x^n$$

for  $x$  close enough to zero, *provided*  $f(x)$  converges to its Taylor series expansion. One can prove convergence by bounding the remainder term in Taylor's theorem. We will omit the details since they are not needed in the formal setting considered here.

Motivated by the formula in the previous example, we define the following series to model the function  $(1+x)^r$ .

**7.69. Definition: Falling Factorials and  $\text{Pow}_r$ .** For every  $r \in K$  and every integer  $n \geq 1$ , define the *falling factorial*

$$(r)_{\downarrow n} = r(r-1)(r-2) \cdots (r-n+1) \in K.$$

Let  $(r)_{\downarrow 0} = 1$ . Define the formal power series

$$\text{Pow}_r = \sum_{n=0}^{\infty} \frac{(r)_{\downarrow n}}{n!} x^n \in K[[x]].$$

(This definition uses the assumption that  $K$  is a field containing  $\mathbb{Q}$ .)

**7.70. Example:  $\text{Pow}_r$  for Integral  $r$ .** Suppose  $r \in \mathbb{N} \subseteq K$ . Let us show that  $\text{Pow}_r = (1+x)^r$  (the product of  $r$  copies of the series  $1+x$ ) in this case. First note that  $(r)_{\downarrow n}$  reduces to the binomial coefficient  $\binom{r}{n} = \frac{r!}{n!(r-n)!}$  when  $r$  is a nonnegative integer (and this coefficient is zero for  $n > r$ ). Next, invoking the binomial theorem 2.14 in the commutative ring  $K[[x]]$ , we see that

$$(1+x)^r = \sum_{n=0}^r \binom{r}{n} x^n = \sum_{n=0}^{\infty} \frac{(r)_{\downarrow n}}{n!} x^n = \text{Pow}_r.$$

Similarly, if  $r = -1$ , the definition of falling factorials shows that  $(r)_{\downarrow n} / n! = (-1)^n$  for all  $n \geq 0$ . On the other hand, we have seen that the multiplicative inverse of  $1+x$  in  $K[[x]]$  is

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + \cdots = \sum_{n=0}^{\infty} (-1)^n x^n = \sum_{n=0}^{\infty} \frac{(r)_{\downarrow n}}{n!} x^n = \text{Pow}_{-1}.$$

So  $\text{Pow}_r = (1+x)^r$  is also true when  $r = -1$ . We will see in a moment that the same result holds for all negative integers  $r$ .

If  $x, r, s$  are real numbers, we have the familiar law of exponents:  $(1+x)^{r+s} = (1+x)^r \cdot (1+x)^s$ . We now prove the formal analogue of this result.

**7.71. Theorem: Formal Exponent Law for  $\text{Pow}_r$ .** For all  $r, s \in K$ ,  $\text{Pow}_{r+s} = \text{Pow}_r \text{Pow}_s$ .

*Proof.* We show  $\text{Pow}_{r+s}(n) = (\text{Pow}_r \text{Pow}_s)(n)$  for each  $n \geq 0$ . On one hand,  $\text{Pow}_{r+s}(n) = (r+s) \downarrow_n / n!$ . On the other hand,

$$(\text{Pow}_r \text{Pow}_s)(n) = \sum_{k=0}^n \text{Pow}_r(k) \text{Pow}_s(n-k) = \sum_{k=0}^n \frac{(r) \downarrow_k}{k!} \frac{(s) \downarrow_{n-k}}{(n-k)!}.$$

Comparing these expressions, we see that we must prove the identities

$$(r+s) \downarrow_n = \sum_{k=0}^n \binom{n}{k} (r) \downarrow_k (s) \downarrow_{n-k}$$

for all  $n \geq 0$ . We use induction on  $n$ . When  $n = 0$ , the identity reads  $1 = \binom{0}{0} \cdot 1 \cdot 1$ , which is true. Assume the identity holds for some  $n \geq 0$ . Using the recursion  $\binom{n+1}{j} = \binom{n}{j-1} + \binom{n}{j}$ , we compute

$$\begin{aligned} (r+s) \downarrow_{n+1} &= (r+s) \downarrow_n (r+s-n) \\ &= \sum_{k=0}^n \binom{n}{k} (r) \downarrow_k (s) \downarrow_{n-k} ((r-k) + (s-(n-k))) \\ &= \sum_{k=0}^n \binom{n}{k} (r) \downarrow_{k+1} (s) \downarrow_{n-k} + \sum_{k=0}^n \binom{n}{k} (r) \downarrow_k (s) \downarrow_{n-k+1} \\ &= \sum_{j=1}^{n+1} \binom{n}{j-1} (r) \downarrow_j (s) \downarrow_{n+1-j} + \sum_{j=0}^n \binom{n}{j} (r) \downarrow_j (s) \downarrow_{n+1-j} \\ &= \sum_{j=0}^{n+1} \binom{n+1}{j} (r) \downarrow_j (s) \downarrow_{n+1-j}. \end{aligned}$$

In the next-to-last step, we changed summation variables to  $j = k+1$  in the first sum, and  $j = k$  in the second sum. The reader should check that the last equality is valid even for the extreme terms  $j = 0$  and  $j = n+1$ . This completes the induction argument.  $\square$

**7.72. Theorem: Negative Binomial Formula.** For every integer  $r > 0$ ,

$$(1+x)^{-r} = \text{Pow}_{-r} = \sum_{n=0}^{\infty} \binom{r+n-1}{n, r-1} (-1)^n x^n \in K[[x]].$$

*Proof.* The first equality follows by iterating 7.71  $r$  times, recalling that  $(1+x)^{-1} = \text{Pow}_{-1}$ . The second equality follows from 7.69 and the identity

$$\begin{aligned} \frac{(-r) \downarrow_n}{n!} &= \frac{(-r)(-r-1)(-r-2) \cdots (-r-(n-1))}{n!} \\ &= (-1)^n \frac{(r+n-1) \downarrow_n}{n!} = \binom{r+n-1}{n} (-1)^n. \quad \square \end{aligned}$$

We have now shown that  $\text{Pow}_r = (1+x)^r$  holds for *all* integers  $r$ . So we can introduce the following notation for  $\text{Pow}_r$  without risk of ambiguity.

**7.73. Definition:**  $(1+x)^r$ . For any  $r \in K$ , let  $(1+x)^r$  denote the series  $\text{Pow}_r \in K[[x]]$ .

## 7.12 Generalized Powers of Formal Series

We now have the necessary tools to define operations such as  $\sqrt{F}$ , provided the formal power series  $F$  satisfies suitable hypotheses.

**7.74. Definition: Generalized Powers.** Suppose  $F \in K[[x]]$  has  $F(0) = 1$ , and  $r \in K$ . Let  $F^r$  be the composition  $\text{Pow}_r \bullet (F - 1)$ , which is defined since  $F - 1$  has zero constant term.

Informally,  $\text{Pow}_r \bullet (F - 1) = (1 + (F - 1))^r$ , so this definition is reasonable. Observe that  $F^r$  always has constant term 1. When  $r = 1/n$  for  $n$  a positive integer, we also write  $\sqrt[n]{F}$  to denote  $F^{1/n}$ .

Many familiar rules for manipulating powers remain true in the formal setting, but they must be reproved formally before they can be used.

**7.75. Theorem: Properties of Formal Powers.** Suppose  $F \in K[[x]]$  has  $F(0) = 1$ . For any  $r, s \in K$ ,  $F^{r+s} = F^r \cdot F^s$ . Furthermore, when  $r$  is an integer,  $F^r$  (as defined in 7.74) coincides with the customary algebraic definition of  $F^r$  (namely the product of  $r$  copies of  $F$  for  $r \geq 0$ , or  $|r|$  copies of  $1/F$  for  $r < 0$ ).

*Proof.* Recall from 7.71 that  $\text{Pow}_{r+s} = \text{Pow}_r \text{Pow}_s$ . Using the fact that composing on the right by  $F - 1$  is a ring homomorphism (see 7.62(b)), we obtain

$$\begin{aligned} F^{r+s} &= \text{Pow}_{r+s} \bullet (F - 1) = (\text{Pow}_r \text{Pow}_s) \bullet (F - 1) \\ &= (\text{Pow}_r \bullet (F - 1))(\text{Pow}_s \bullet (F - 1)) = F^r F^s. \end{aligned}$$

For the rest of this proof,  $n$  will denote an integer and  $F^n$  will have the usual algebraic meaning (repeated multiplication). We must prove  $\text{Pow}_n \bullet (F - 1) = F^n$  for all  $n \in \mathbb{Z}$ . When  $n = 0$ ,  $\text{Pow}_0 \bullet (F - 1) = 1 \bullet (F - 1) = 1 = F^0$ . When  $n = 1$ ,  $\text{Pow}_1 \bullet (F - 1) = (1 + x) \bullet (F - 1) = 1 + (F - 1) = F = F^1$ . By induction, assuming  $n \geq 1$  and  $\text{Pow}_n \bullet (F - 1) = F^n$ , the result just proved shows that

$$\text{Pow}_{n+1} \bullet (F - 1) = (\text{Pow}_n \bullet (F - 1))(\text{Pow}_1 \bullet (F - 1)) = F^n F^1 = F^{n+1}.$$

(The last step uses the algebraic definition of the power  $F^{n+1}$ .) Similarly, the known identity

$$(\text{Pow}_{-n} \bullet (F - 1))(\text{Pow}_n \bullet (F - 1)) = \text{Pow}_0 \bullet (F - 1) = 1$$

shows that  $\text{Pow}_{-n} \bullet (F - 1)$  is the multiplicative inverse of  $\text{Pow}_n \bullet (F - 1) = F^n$ . In other words,  $\text{Pow}_{-n} \bullet (F - 1) = (F^n)^{-1} = F^{-n}$ .  $\square$

**7.76. Example: Negative Binomial Expansion.** Suppose  $F = 1 - cx$  where  $c \in K$  is a constant. Let  $r$  be a positive integer. Using 7.72 and the definition of composition, we find that

$$(1 - cx)^{-r} = \left( \frac{1}{1 - cx} \right)^r = \sum_{n=0}^{\infty} \binom{n+r-1}{n, r-1} c^n x^n. \quad (7.7)$$

This identity is used often when computing with generating functions.

Next we prove a partial version of another familiar law of exponents.

**7.77. Theorem: Iterated Exponents.** For all  $F \in K[[x]]$ , all  $r \in K$  and all integers  $n$ ,  $(F^r)^n = F^{rn}$ .

*Proof.* The idea is to iterate the known identity  $F^{r+s} = F^r F^s$ , which holds for all  $r, s \in K$ . First, we prove the result for integers  $n \geq 0$  by induction. When  $n = 0$ ,  $(F^r)^0 = 1 = F^{r0}$ . When  $n = 1$ ,  $(F^r)^1 = F^r = F^{r1}$ . Assuming  $n \geq 1$  and  $(F^r)^n = F^{rn}$  is already known, we calculate

$$(F^r)^{n+1} = (F^r)^n (F^r)^1 = F^{rn} F^r = F^{rn+r} = F^{r(n+1)}.$$

Next,  $(F^r)^{-n} (F^r)^n = (F^r)^{-n+n} = 1$ , and hence  $(F^r)^{-n} = ((F^r)^n)^{-1} = (F^{rn})^{-1}$ . Similarly,  $F^{-rn} F^{rn} = F^{-rn+rn} = 1$ , so that  $(F^{rn})^{-1} = F^{-rn}$ . So finally  $(F^r)^{-n} = F^{r(-n)}$ , which establishes the result for negative integers.  $\square$

The next result is the analogue of the fact that every positive real number has a unique positive  $n$ th root, for each  $n \geq 0$ .

**7.78. Theorem: Existence and Uniqueness of  $n$ th Roots.** Suppose  $F \in K[[x]]$  satisfies  $F(0) = 1$ . For every integer  $n \geq 1$ , there exists a unique  $G \in K[[x]]$  with  $G(0) = 1$  such that  $G^n = F$ , namely  $G = F^{1/n} = \sqrt[n]{F}$ .

*Proof.* Existence of  $G$  follows from the previous result, since  $(F^{1/n})^n = F^{(1/n) \cdot n} = F^1 = F$ . To prove uniqueness, suppose  $G, H \in K[[x]]$  satisfy  $G^n = F = H^n$  and  $G(0) = 1 = H(0)$ . By the distributive law, we have the factorization

$$0 = G^n - H^n = (G - H)(G^{n-1} + G^{n-2}H^1 + G^{n-3}H^2 + \cdots + H^{n-1}).$$

Since  $K[[x]]$  is an integral domain, this implies that either  $G - H = 0$  or  $\sum_{i=0}^{n-1} G^{n-1-i} H^i = 0$ . The first alternative gives  $G = H$ , as desired. The second alternative is impossible, since the left side has constant term  $\sum_{i=0}^{n-1} 1^{n-1-i} 1^i = n > 0$  while the right side has constant term zero. (This proof uses the assumption that  $K$  is a field containing  $\mathbb{Q}$ .)  $\square$

**7.79. Theorem: Formal Power Rule.** Suppose  $F \in K[[x]]$  satisfies  $F(0) = 1$ . For  $r \in K$ ,

$$(F^r)' = r F^{r-1} F'.$$

*Proof.* Let us first show that  $\text{Pow}'_r = r \text{Pow}_{r-1}$ . The formal derivative of  $\text{Pow}_r = \sum_{n \geq 0} ((r) \downarrow_n / n!) x^n$  is

$$\text{Pow}'_r = \sum_{n \geq 0} \frac{(n+1)(r) \downarrow_{n+1}}{(n+1)!} x^n = \sum_{n \geq 0} \frac{r(r-1) \downarrow_n}{n!} x^n = r \text{Pow}_{r-1}.$$

It follows that

$$(F^r)' = [\text{Pow}_r \bullet (F-1)]' = [\text{Pow}'_r \bullet (F-1)](F-1)' = [r \text{Pow}'_{r-1} \bullet (F-1)]F' = r F^{r-1} F'. \quad \square$$

We now have the necessary machinery to solve certain quadratic equations involving formal power series.

**7.80. Example.** Suppose  $F \in \mathbb{Q}[[x]]$  is a formal series such that  $xF^2 - F + 1 = 0$ . Let us “solve for  $F$ ,” determining  $F_n$  for all  $n \in \mathbb{N}$ . The given equation immediately implies  $F_0 = 1$ . Multiplying the quadratic equation by  $4x$  gives  $4x^2 F^2 - 4xF + 4x = 0$ . Completing the square leads to  $(1 - 2xF)^2 = 1 - 4x$ . Since  $\sqrt{1 - 4x}$  is the *unique* power series with constant term 1 that squares to  $1 - 4x$ , we conclude that

$$1 - 2xF = \sqrt{1 - 4x}.$$

Rearranging,  $xF = \frac{1}{2}(1 - \sqrt{1 - 4x})$ . Since the power series on the right has zero constant term, we may safely write

$$F = \frac{1 - \sqrt{1 - 4x}}{2x},$$

where the notation  $\frac{1}{x} \sum_{n \geq 1} a_n x^n$  can be regarded as shorthand for the series  $\sum_{n \geq 0} a_{n+1} x^n$ . (Note  $x$  is not a unit in  $\mathbb{Q}[[x]]$ , so algebraic division by  $x$  is not permitted in this ring, although we could allow it by passage to the field of fractions (§7.7). What we are really doing is cancelling the nonzero element  $x$  in the integral domain  $\mathbb{Q}[[x]]$ ; cf. 7.135.)

We remark that our formula for  $F$  is exactly what we would have obtained by formally applying the quadratic formula for solving  $AF^2 + BF + C = 0$ , which gives  $F = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$ . However, one must take care in blindly applying this formula since we cannot divide by arbitrary formal power series, and the sign ambiguity in the square root must be resolved somehow. In our example, we are forced to choose the minus sign.

Now we are ready to find  $F_n$  for each  $n \geq 0$ . By 7.74,

$$\sqrt{1-4x} = \sum_{m=0}^{\infty} \frac{(1/2) \downarrow_m}{m!} (-4x)^m.$$

For  $m \geq 1$ , the coefficient of  $x^m$  here is

$$\begin{aligned} (-1)^m 4^m \frac{(1/2)(-1/2)(-3/2) \cdots (-(2m-3)/2)}{m!} &= -\frac{2^m \cdot 1 \cdot 1 \cdot 3 \cdot 5 \cdots (2m-3)}{m!} \\ &= -\frac{2^m (2m-2)!}{m! \cdot 2 \cdot 4 \cdot 6 \cdots (2m-2)} \\ &= -2 \frac{(2m-2)!}{m!(m-1)!}, \end{aligned}$$

where the last step follows by using  $m-1$  powers of 2 in the numerator to divide each of the even numbers in the denominator by 2. It now follows that

$$\left( \frac{1 - \sqrt{1-4x}}{2} \right)_m = \frac{1}{2m-1} \binom{2m-1}{m, m-1} \quad (m \geq 1).$$

Finally,

$$F_n = \left( \frac{1 - \sqrt{1-4x}}{2} \right)_{n+1} = \frac{1}{2n+1} \binom{2n+1}{n+1, n} \quad (n \geq 0).$$

Thus,  $F = \sum_{n \geq 0} C_n x^n$  is the generating function for the Catalan numbers (see 1.55).

Calculations such as those in the preceding example occur frequently in the application of formal power series to combinatorial problems.

## 7.13 Partial Fraction Expansions

Suppose  $g$  is a polynomial in  $\mathbb{C}[x]$  with nonzero constant term, and  $f$  is any polynomial in  $\mathbb{C}[x]$ . We have seen (§7.6) that  $g$  is a unit in  $\mathbb{C}[[x]]$ , so that we can write  $f/g = \sum_{n=0}^{\infty} b_n x^n$  for suitable complex numbers  $b_n$ . This section presents a technique for finding explicit expressions for the coefficients  $b_n$ , which is a formal version of the “method of partial fractions” from calculus. We will see that this technique can be used to find explicit closed formulas for certain recursively defined sequences. Our starting point is the famous fundamental theorem of algebra, which we state here without proof.

**7.81. Fundamental Theorem of Algebra.** Let  $p \in \mathbb{C}[x]$  be a monic polynomial of degree  $n \geq 1$ . There exist pairwise distinct complex numbers  $r_1, \dots, r_k$  (unique up to reordering) and unique positive integers  $n_1, \dots, n_k$  such that

$$p = (x - r_1)^{n_1} (x - r_2)^{n_2} \cdots (x - r_k)^{n_k} \in \mathbb{C}[x].$$

The number  $r_i$  is called a *root* of  $p$  of *multiplicity*  $n_i$ .

The following variant of the fundamental theorem is needed in partial fraction problems because of the form of the negative binomial expansion (see 7.76).

**7.82. Theorem: Factorization of Polynomials in  $\mathbb{C}[x]$ .** Let  $p \in \mathbb{C}[x]$  be a polynomial of degree  $n \geq 1$  with  $p(0) = 1$ . There exist pairwise distinct, nonzero complex numbers  $r_1, \dots, r_k$  and positive integers  $n_1, \dots, n_k$  such that

$$p(x) = (1 - r_1 x)^{n_1} (1 - r_2 x)^{n_2} \cdots (1 - r_k x)^{n_k} \in \mathbb{C}[x].$$

*Proof.* Consider the polynomial  $q = x^n P_p(1/x)$ . We have  $p = \sum_{i=0}^n p_i x^i$  and  $q = \sum_{i=0}^n p_{n-i} x^i$ , so that  $q$  is obtained from  $p$  by “reversing the coefficient sequence.” Since  $p_0 = 1$ ,  $q$  is a monic polynomial of degree  $n$ . Using the fundamental theorem of algebra, we write

$$x^n P_p(1/x) = q = \prod_{i=1}^k (x - r_i)^{n_i},$$

where  $\sum n_i = n$ . Since the constant term of  $q$  is nonzero, no  $r_i$  is equal to zero. Reversing the coefficient sequence again, it follows that

$$p = x^n P_q(1/x) = x^n \prod_{i=1}^k ((1/x) - r_i)^{n_i} = \prod_{i=1}^k x^{n_i} \left( \frac{1 - r_i x}{x} \right)^{n_i} = \prod_{i=1}^k (1 - r_i x)^{n_i}. \quad \square$$

The next step is to rewrite a general fraction  $f/g$  as a sum of fractions whose denominators have the form  $(1 - rx)^m$ . Note that, as long as  $g(0) \neq 0$ , we can always arrange  $g(0) = 1$  by multiplying numerator and denominator by a suitable scalar in  $K$ .

**7.83. Theorem: Splitting a Denominator.** Suppose  $f, g \in \mathbb{C}[x]$  are polynomials such that  $g(0) = 1$ , and let  $g$  have factorization  $g(x) = \prod_{i=1}^k (1 - r_i x)^{n_i}$ , where  $r_1, \dots, r_k \in \mathbb{C}$  are distinct and nonzero. There exist polynomials  $p_0, p_1, \dots, p_k$  with  $\deg(p_i) < n_i$  (or  $p_i = 0$ ) for  $1 \leq i \leq k$ , such that

$$\frac{f}{g} = p_0 + \sum_{i=1}^k \frac{p_i}{(1 - r_i x)^{n_i}}.$$

*Proof.* For  $1 \leq i \leq k$ , define a polynomial  $h_i = g/(1 - r_i x)^{n_i} = \prod_{j:j \neq i} (1 - r_j x)^{n_j}$ . Since  $r_1, \dots, r_k$  are distinct,  $\gcd(h_1, \dots, h_k) = 1$ . By a well-known result from polynomial algebra, it follows that there exist polynomials  $q_1, \dots, q_k \in \mathbb{C}[x]$  with  $q_1 h_1 + \cdots + q_k h_k = 1$ . Therefore,

$$\begin{aligned} \frac{f}{g} &= \frac{f \cdot 1}{g} = \frac{f q_1 h_1 + \cdots + f q_k h_k}{g} \\ &= \sum_{i=1}^k \frac{f q_i}{(1 - r_i x)^{n_i}}. \end{aligned}$$

This is almost the answer we want, but the degrees of the numerators may be too high. Using

polynomial division (see 5.87), we can write  $f q_i = a_i(1 - r_i x)^{n_i} + p_i$  where  $a_i, p_i \in \mathbb{C}[x]$ , and either  $p_i = 0$  or  $\deg(p_i) < n_i$ . Dividing by  $(1 - r_i x)^{n_i}$ , we see that

$$\frac{f}{g} = p_0 + \sum_{i=1}^k \frac{p_i}{(1 - r_i x)^{n_i}}$$

holds if we take  $p_0 = \sum_{i=1}^k a_i \in \mathbb{C}[x]$ .  $\square$

The fractions  $p_i/(1 - r_i x)^{n_i}$  (with  $\deg(p_i) < n_i$  or  $p_i = 0$ ) can be further reduced into sums of fractions where the numerators are complex constants.

**7.84. Theorem: Division by  $(1 - rx)^n$ .** Given a fraction  $p/(1 - rx)^n$  where  $p \in \mathbb{C}[x]$ ,  $\deg(p) < n$  (or  $p = 0$ ), and  $0 \neq r \in \mathbb{C}$ , there exist complex numbers  $a_1, \dots, a_n$  such that

$$\frac{p}{(1 - rx)^n} = \sum_{i=1}^n \frac{a_i}{(1 - rx)^i}.$$

*Proof.* Consider the evaluation homomorphism  $E : \mathbb{C}[x] \rightarrow \mathbb{C}[x]$  such that  $E(x) = 1 - rx$  (see 7.22). The evaluation homomorphism  $E' : \mathbb{C}[x] \rightarrow \mathbb{C}[x]$  such that  $E'(x) = (1 - x)/r$  is a two-sided inverse to  $E$  (since  $E(E'(x)) = x = E'(E(x))$ ), so  $E$  is a bijection. In particular,  $E$  is surjective, so  $p = E(q)$  for some  $q \in \mathbb{C}[x]$ . Now, one may check that  $E$  and  $E'$  each map polynomials of degree  $< n$  to polynomials of degree  $< n$ , and it follows that  $\deg(q) < n$  (or  $q = 0$ ). Write  $q = c_0 + c_1 x + c_2 x^2 + \dots + c_{n-1} x^{n-1}$ , with  $c_i \in \mathbb{C}$ . Then

$$p = E(q) = c_0 + c_1(1 - rx) + c_2(1 - rx)^2 + \dots + c_{n-1}(1 - rx)^{n-1}.$$

Dividing by  $(1 - rx)^n$ , we see that we may take  $a_1 = c_{n-1}, \dots, a_{n-1} = c_1, a_n = c_0$ .  $\square$

The next result summarizes the partial fraction manipulations in the last two theorems. The uniqueness proof given below also provides a convenient algorithm for finding the coefficients in the partial fraction decomposition.

**7.85. Theorem: Partial Fraction Decompositions in  $\mathbb{C}(x)$ .** Suppose  $f, g \in \mathbb{C}[x]$  are polynomials with  $g(0) = 1$ ; let  $g = \prod_{i=1}^k (1 - r_i x)^{n_i}$  where the  $r_i$  are distinct nonzero complex numbers. There exist a unique polynomial  $h \in \mathbb{C}[x]$  and unique complex numbers  $a_{ij}$  (where  $1 \leq i \leq k, 1 \leq j \leq n_i$ ) with

$$\frac{f}{g} = h + \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{a_{ij}}{(1 - r_i x)^j}. \quad (7.8)$$

Viewing  $f/g \in \mathbb{C}[[x]]$ , we have (for all  $m \in \mathbb{N}$ )

$$(f/g)_m = h_m + \sum_{i=1}^k \sum_{j=1}^{n_i} a_{ij} \binom{m+j-1}{m, j-1} r_i^m.$$

*Proof.* Existence of the decomposition follows by combining 7.83 and 7.84. The formula for the coefficient of  $x^m$  follows from the negative binomial expansion 7.76. We must now prove uniqueness of  $h$  and the  $a_{ij}$ 's. Note first that the numbers  $r_i$  and  $n_i$  appearing in the factorization of  $g$  are unique (this follows from the uniqueness assertion in the fundamental theorem of algebra). Now consider any expression of the form (7.8). Multiplying both sides by  $g$  produces an equation

$$f = gh + \sum_{i=1}^k \sum_{j=1}^{n_i} a_{ij} (1 - r_i x)^{n_i-j} \prod_{s \neq i} (1 - r_s x)^{n_s}, \quad (7.9)$$

where both sides are polynomials. Furthermore, the terms in the double sum add up to a polynomial that is either zero or has degree less than  $\deg(g)$ . Thus  $h$  must be the quotient when  $f$  is divided by  $g$  using the polynomial division algorithm, and this quotient is known to be unique. Next, we show how to recover the “top coefficients”  $a_{i,n_i}$  for  $1 \leq i \leq k$ . Fix  $i$ , and apply the functions associated to the polynomials on each side of (7.9) to  $z = 1/r_i \in \mathbb{C}$ . Since any positive power of  $(1 - r_i x)$  becomes zero for this choice of  $x$ , all but one term on the right side becomes zero. We are left with

$$P_f(1/r_i) = a_{i,n_i} \prod_{s \neq i} (1 - r_s/r_i)^{n_s}.$$

Since  $r_s \neq r_i$  for  $s \neq i$ , the product is nonzero. Thus there is a unique  $a_{i,n_i} \in \mathbb{C}$  for which this equation holds. We can use the displayed formula to calculate each  $a_{i,n_i}$  given  $f$  and  $g$ .

To find the remaining  $a_{ij}$ 's, subtract the recovered summands  $a_{i,n_i}/(1 - r_i x)^{n_i}$  from both sides of (7.8) (thus replacing  $f/g$  by a new fraction  $f_1/g_1$ ) to obtain a new problem in which all  $n_i$ 's have been reduced by one. We now repeat the procedure of the previous paragraph to find  $a_{i,n_i-1}$  for all  $i$ . Continuing similarly, we eventually recover all the  $a_{ij}$ .  $\square$

**7.86. Example.** Let us find the partial fraction expansion of

$$\frac{f}{g} = \frac{x^2 - 2}{1 - 2x - x^2 + 2x^3}.$$

To find the required factorization of the denominator, we first reverse the coefficient sequence to obtain  $x^3 - 2x^2 - x + 2$ . This polynomial factors as  $(x - 2)(x - 1)(x + 1)$ , so the original denominator can be rewritten as

$$1 - 2x - x^2 + 2x^3 = (1 - 2x)(1 - x)(1 + x)$$

(see the proof of 7.82). We know that

$$\frac{x^2 - 2}{1 - 2x - x^2 + 2x^3} = \frac{A}{1 - 2x} + \frac{B}{1 - x} + \frac{C}{1 + x} \quad (7.10)$$

for suitable complex constants  $A, B, C$ . To find  $A$ , multiply both sides by  $1 - 2x$  to get

$$\frac{x^2 - 2}{(1 - x)(1 + x)} = A + \frac{B(1 - 2x)}{1 - x} + \frac{C(1 - 2x)}{1 + x}.$$

Now set  $x = 1/2$  to see that  $A = (-7/4)/(3/4) = -7/3$ . Similarly,

$$\begin{aligned} B &= \left. \frac{x^2 - 2}{(1 - 2x)(1 + x)} \right|_{x=1} = 1/2; \\ C &= \left. \frac{x^2 - 2}{(1 - 2x)(1 - x)} \right|_{x=-1} = -1/6. \end{aligned}$$

It now follows from (7.10) that

$$\left( \frac{x^2 - 2}{1 - 2x - x^2 + 2x^3} \right)_n = -\frac{7}{3} \cdot 2^n + \frac{1}{2} - \frac{1}{6} \cdot (-1)^n \quad (n \in \mathbb{N}).$$

**7.87. Example.** We will find the partial fraction expansion of

$$\frac{f}{g} = \frac{1}{1 - 9x + 30x^2 - 46x^3 + 33x^4 - 9x^5}.$$



Factoring the numerator as in the last example, we find that  $g(x) = (1-x)^3(1-3x)^2$ . We can therefore write

$$\frac{f}{g} = \frac{A}{(1-x)^3} + \frac{B}{(1-x)^2} + \frac{C}{1-x} + \frac{D}{(1-3x)^2} + \frac{E}{1-3x}. \quad (7.11)$$

To find  $A$ , multiply both sides by  $(1-x)^3$  and then substitute  $x = 1$  to get  $A = 1/(-2)^2 = 1/4$ . Similarly, multiplication by  $(1-3x)^2$  reveals that  $D = 1/(2/3)^3 = 27/8$ . Having found  $A$  and  $D$ , we subtract  $A/(1-x)^3$  and  $D/(1-3x)^2$  from both sides of (7.11). After simplifying, we are left with

$$\frac{(3/8)(3x-7)}{(1-x)^2(1-3x)} = \frac{B}{(1-x)^2} + \frac{C}{1-x} + \frac{E}{1-3x}.$$

Now we repeat the process. Multiplying by  $(1-x)^2$  and setting  $x = 1$  shows that  $B = 3/4$ . Similarly,  $E = -81/16$ . Subtracting these terms from both sides leaves  $(27/16)/(1-x)$ , so  $C = 27/16$ . Using (7.11) and (7.76), we conclude that

$$(f/g)_n = \frac{1}{4} \binom{n+2}{2} + \frac{3}{4} \binom{n+1}{1} + \frac{27}{16} + \frac{27}{8} \binom{n+1}{1} 3^n - \frac{81}{16} 3^n \quad (n \in \mathbb{N}).$$

## 7.14 Application to Recursions

In Chapter 2, we saw that many enumeration problems in combinatorics lead naturally to recursion relations. Formal power series and partial fraction expansions provide a powerful method for solving a wide class of recursions. Before stating the general method, we consider some typical examples.

**7.88. Example.** In 2.22, we found that the number  $a_n$  of subsets of an  $n$ -element set satisfies the following recursion and initial condition:

$$a_n = 2a_{n-1} \quad (n \geq 1); \quad a_0 = 1.$$

It is not hard to guess that the solution to this recursion is  $a_n = 2^n$  for all  $n \geq 0$ , and it is then routine to prove that this guess is correct by induction on  $n$ . However, for more complicated recursions, one is unlikely to find the solution by guessing. Thus, let us see how to solve the recursion using formal power series.

We introduce the formal series  $F = \sum_{n \geq 0} a_n x^n$  whose coefficients are given by the unknown sequence  $(a_n)$ . Notice that  $xF = \sum_{m \geq 0} a_m x^{m+1} = \sum_{n \geq 1} a_{n-1} x^n$ . By the recursive description of the  $a_n$ 's, we see that  $(F - 2xF)_n = 0$  for all  $n \geq 1$ . On the other hand,  $(F - 2xF)_0 = F_0 = a_0 = 1$  by the initial condition. It follows that

$$F - 2xF = 1 \in \mathbb{C}[[x]].$$

Solving for  $F$ , we find that

$$F = \frac{1}{1-2x} = \sum_{n=0}^{\infty} 2^n x^n.$$

Comparing coefficients of  $x^n$  leads to the expected solution  $a_n = 2^n$ .

Now let us modify the problem by changing the initial condition to  $a_0 = 3$ . The same reasoning as above leads to  $F = 3/(1-2x)$ , so that the new solution is  $a_n = 3 \cdot 2^n$ .

For a more subtle modification, let us change the recursion to  $a_n = 2a_{n-1} + 1$ , with

initial condition  $a_0 = 0$ . (This recursion describes the number of moves needed to solve the famous “Tower of Hanoi” puzzle.) Define  $F = \sum_{n \geq 0} a_n x^n$  as before. We now have  $(F - 2xF)_n = a_n - 2a_{n-1} = 1$  for all  $n \geq 1$ , and  $(F - 2xF)_0 = a_0 = 0$ . We conclude that

$$F - 2xF = x + x^2 + x^3 + \cdots + x^n + \cdots = \frac{x}{1-x}.$$

Solving for  $F$  and using partial fractions, we get

$$F = \frac{x}{(1-x)(1-2x)} = \frac{1}{1-2x} - \frac{1}{1-x}.$$

Extracting the coefficient of  $x^n$  yields the solution  $a_n = 2^n - 1^n = 2^n - 1$  for all  $n \geq 0$ .

**7.89. Example: Fibonacci Recursion.** The *Fibonacci numbers* are defined by the recursion  $f_n = f_{n-1} + f_{n-2}$  for all  $n \geq 2$ , with initial conditions  $f_0 = 0$ ,  $f_1 = 1$ . (Sometimes other initial conditions are used, such as  $f_0 = f_1 = 1$ , which leads to a shift in the indexing of the sequence.) Let us use formal power series to find an explicit closed formula for the numbers  $f_n$ . Define  $F = \sum_{n \geq 0} f_n x^n$ . Since  $(xF)_n = f_{n-1}$  and  $(x^2F)_n = f_{n-2}$  for all  $n \geq 2$ , the recursion gives

$$(F - xF - x^2F)_n = 0 \quad (n \geq 2).$$

On the other hand, the initial conditions show that

$$(F - xF - x^2F)_0 = f_0 = 0; \quad (F - xF - x^2F)_1 = f_1 - f_0 = 1.$$

It follows that  $F - xF - x^2F = 0 + 1x + 0x^2 + \cdots = x$ . Solving for  $F$  gives

$$F = \frac{x}{1-x-x^2}.$$

We now apply the method of partial fractions. First, reversing the coefficient sequence in the denominator gives the polynomial  $x^2 - x - 1 = (x - r_1)(x - r_2)$ , where (by the quadratic formula)

$$r_1 = \frac{1 + \sqrt{5}}{2}, \quad r_2 = \frac{1 - \sqrt{5}}{2}.$$

It follows that  $1 - x - x^2 = (1 - r_1x)(1 - r_2x)$ . Next write

$$F = \frac{x}{(1 - r_1x)(1 - r_2x)} = \frac{A}{1 - r_1x} + \frac{B}{1 - r_2x}.$$

Multiplying both sides by  $(1 - r_1x)$  and setting  $x = 1/r_1$ , we find that  $A = (1/r_1)/(1 - r_2/r_1) = 1/(r_1 - r_2) = 1/\sqrt{5}$ . Similarly we find that  $B = -1/\sqrt{5}$ , so that

$$F = \frac{x}{(1 - r_1x)(1 - r_2x)} = \frac{1}{\sqrt{5}} \left( \frac{1}{1 - r_1x} - \frac{1}{1 - r_2x} \right).$$

Extracting the coefficient of  $x^n$ , we conclude that the Fibonacci numbers are given by the following exact formula:

$$f_n = (r_1^n - r_2^n)/\sqrt{5} = \frac{1}{2^n \sqrt{5}} \left[ (1 + \sqrt{5})^n - (1 - \sqrt{5})^n \right] \quad (n \geq 0).$$

Note that  $|r_2| \approx 0.618 < 1$ , so that  $\lim_{n \rightarrow \infty} r_2^n = 0$ . It follows that, for very large  $n$ ,

$$f_n \approx r_1^n / \sqrt{5} \approx (0.447214) \cdot (1.61803)^n.$$

This formula tells us the asymptotic growth rate of the Fibonacci numbers.

The next theorem gives a general method for solving recursion relations with constant coefficients.

**7.90. Theorem: Recursions with Constant Coefficients.** Suppose we are given the following data: a positive integer  $k$ , constants  $c_1, c_2, \dots, c_k, d_0, \dots, d_{k-1} \in K$ , and a function  $g: \mathbb{N} \rightarrow K$ . The recursion

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \dots + c_k a_{n-k} + g(n) \quad (n \geq k)$$

with initial conditions  $a_i = d_i$  for  $0 \leq i < k$  has a unique solution. Setting  $d'_i = d_i - c_1 d_{i-1} - c_2 d_{i-2} - \dots - c_i d_0$ ,  $F = \sum_{n \geq 0} a_n x^n$ ,  $G = \sum_{i=0}^{k-1} d'_i x^i + \sum_{n \geq k} g(n) x^n$ , and  $p = 1 - c_1 x - c_2 x^2 - \dots - c_k x^k$ , we have  $F = G/p$ .

*Proof.* The existence and uniqueness of the sequence  $(a_n : n \geq 0)$  satisfying the given recursion and initial conditions is intuitively plausible and can be informally established by an induction argument. (A formal proof requires the *recursion theorem* from set theory; see Section 12 of Halmos [66] for a discussion of this theorem.) It follows that the formal series  $F \in K[[x]]$  in the theorem statement is well defined. Consider next the formal series

$$H = (1 - c_1 x - c_2 x^2 - \dots - c_k x^k) F = pF.$$

For each  $n \geq k$ , the recursion shows that

$$H_n = a_n - c_1 a_{n-1} - \dots - c_k a_{n-k} = g(n) = G_n.$$

On the other hand, for  $0 \leq n < k$ , the initial conditions show that

$$H_n = d'_n = G_n.$$

So  $H = G$ , and the formula for  $F$  follows by dividing the equation  $G = pF$  by the invertible element  $p$ .  $\square$

**7.91. Example.** Let us solve the recursion

$$a_n = 5a_{n-1} - 6a_{n-2} + 2^n \quad (n \geq 2)$$

subject to the initial conditions  $a_0 = 0, a_1 = 1$ . Use the theorem with  $k = 2, c_1 = 5, c_2 = -6, d_0 = 0, d_1 = 1$ , and  $g(n) = 2^n$ . We find that  $d'_0 = 0, d'_1 = 1, G = 0 + 1x + \sum_{n \geq 2} 2^n x^n = (1 - 2x)^{-1} - 1 - x, p = 1 - 5x + 6x^2 = (1 - 2x)(1 - 3x)$ , and finally

$$F = \frac{(1 - 2x)^{-1} - 1 - x}{(1 - 2x)(1 - 3x)} = \frac{x + 2x^2}{(1 - 2x)^2(1 - 3x)}.$$

A tedious but routine partial fraction computation gives us

$$F = -2(1 - 2x)^{-2} - 3(1 - 2x)^{-1} + 5(1 - 3x)^{-1}.$$

Finally, using 7.76, we obtain

$$a_n = F_n = -2 \binom{n+1}{1} 2^n - 3 \cdot 2^n + 5 \cdot 3^n = 5 \cdot 3^n - (2n+5) \cdot 2^n.$$

Once this formula has been found, one may check the answer by proving by induction that it satisfies the initial condition and recursion.

## 7.15 Formal Exponentiation and Formal Logarithms

In 7.67, we introduced formal series

$$E = \sum_{n \geq 1} x^n/n! = e^x - 1, \quad L = \sum_{n \geq 1} (-1)^{n-1} x^n/n = \log(1+x),$$

and showed that  $L \bullet E = x = E \bullet L$ . We can use these series and formal composition to define the exponential and logarithm of certain power series.

**7.92. Definition:**  $\exp(G)$  and  $\log(1+G)$ . Suppose  $G$  is a formal power series with constant term zero. We define

$$e^G = \exp(G) = (E \bullet G) + 1 = \sum_{n=0}^{\infty} G^n/n!; \quad \log(1+G) = L \bullet G = \sum_{n=1}^{\infty} (-1)^{n-1} G^n/n.$$

Similarly, if  $H$  is a formal power series with  $H(0) = 1$ , define

$$\log H = \log(1 + [H - 1]) = L \bullet (H - 1).$$

The combinatorial significance of exponentiating a formal power series will be revealed in the next chapter (see 8.32). For the moment, we will be content to prove some properties of the ordinary exponential and logarithm functions that are also satisfied by their formal counterparts.

**7.93. Theorem: Sum-to-Product Rule for Exponentials.** For all  $G, H \in K[[x]]$  with  $G(0) = 0 = H(0)$ , we have

$$\exp(G+H) = \exp(G)\exp(H).$$

More generally, given a sequence  $G_k \in K[[x]]$  with  $G_k(0) = 0$  for all  $k$  and  $\lim_{k \rightarrow \infty} G_k = 0$ ,

$$\exp\left(\sum_{k=1}^{\infty} G_k\right) = \prod_{k=1}^{\infty} \exp(G_k).$$

*Proof.* To prove the first identity, look at the coefficient of  $x^n$  on each side. For the left side, the binomial theorem gives

$$\begin{aligned} \exp(G+H)_n &= \left( \sum_{k=0}^{\infty} \frac{(G+H)^k}{k!} \right)_n = \left( \sum_{k=0}^n \frac{(G+H)^k}{k!} \right)_n \\ &= \left( \sum_{k=0}^n \sum_{j=0}^k \binom{k}{j} \frac{G^j H^{k-j}}{k!} \right)_n = \sum_{(i,j) \in \mathbb{N}^2: i+j \leq n} \left( \frac{G^j H^i}{j! i!} \right)_n \\ &= \sum_{(i,j) \in \mathbb{N}^2: i+j \leq n} \sum_{m=0}^n \frac{G^j(m)}{j!} \cdot \frac{H^i(n-m)}{i!}. \end{aligned}$$

On the right side, we get

$$\begin{aligned}
 [\exp(G) \exp(H)]_n &= \sum_{m=0}^n \exp(G)_m \exp(H)_{n-m} \\
 &= \sum_{m=0}^n \left( \sum_{j \geq 0} \frac{G^j(m)}{j!} \right) \left( \sum_{i \geq 0} \frac{H^i(n-m)}{i!} \right) \\
 &= \sum_{i,j,m=0}^n \frac{G^j(m)}{j!} \cdot \frac{H^i(n-m)}{i!}.
 \end{aligned}$$

The two answers almost agree, but the ranges of summation for  $i$  and  $j$  do not quite match. However, consider a triple  $(i, j, m)$  in the last summation for which  $i + j > n$ . This forces either  $j > m$  or  $i > n - m$ , so either  $G^j(m) = 0$  or  $H^i(n - m) = 0$ . In any case, the summand indexed by this triple  $(i, j, m)$  is zero. Dropping these summands, we get precisely the sum occurring in the earlier calculation.

Iteration of the result just proved shows that  $\exp\left(\sum_{k=1}^N G_k\right) = \prod_{k=1}^N \exp(G_k)$  for any (finite)  $N \in \mathbb{N}$ . To prove the same formula with  $N = \infty$ , we check the coefficient of  $x^M$  on each side. One sees immediately from the definition that  $\text{ord}(F) > M$  implies  $\text{ord}(\exp(F) - 1) > M$ . Choose  $k_0$  large enough that  $\text{ord}(G_k) > M$  or  $G_k = 0$  for all  $k > k_0$ . Taking  $F = \sum_{k > k_0} G_k$ , we then have  $\text{ord}(F) > M$  or  $F = 0$ . Write  $\exp(F) = 1 + H$  where  $\text{ord}(H) > M$  or  $H = 0$ . Using the result for finite sums gives

$$\begin{aligned}
 \exp\left(\sum_{k=1}^{\infty} G_k\right)_M &= \exp\left(\sum_{k=1}^{k_0} G_k + F\right)_M \\
 &= \left[\exp(F) \prod_{k=1}^{k_0} \exp(G_k)\right]_M = \left[\prod_{k=1}^{k_0} \exp(G_k)\right]_M.
 \end{aligned}$$

Now, for any  $k_1 \geq k_0$ ,

$$\left[\prod_{k=1}^{k_0} \exp(G_k)\right]_M = \left[\prod_{k=1}^{k_1} \exp(G_k)\right]_M$$

since, for  $k_0 < k \leq k_1$ ,  $\exp(G_k)$  is 1 plus terms of order larger than  $M$ . We conclude finally that

$$\exp\left(\sum_{k=1}^{\infty} G_k\right)_M = \left[\prod_{k=1}^{\infty} \exp(G_k)\right]_M$$

for every  $M \geq 0$ . □

**7.94. Theorem: Exponential and Logarithm are Inverses.** If  $H \in K[[x]]$  satisfies  $H(0) = 1$ , then  $\exp(\log(H)) = H$ . If  $G \in K[[x]]$  satisfies  $G(0) = 0$ , then  $\log(\exp(G)) = G$ .

*Proof.* Recall from 7.67 that  $E \bullet L = x = L \bullet E$ . We can therefore compute

$$\begin{aligned}
 \exp(\log(H)) &= (E \bullet (L \bullet (H - 1))) + 1 = ((E \bullet L) \bullet (H - 1)) + 1 \\
 &= (x \bullet (H - 1)) + 1 = H - 1 + 1 = H;
 \end{aligned}$$

$$\begin{aligned}
 \log(\exp(G)) &= L \bullet (([E \bullet G] + 1) - 1) = L \bullet (E \bullet G) \\
 &= (L \bullet E) \bullet G = x \bullet G = G. \quad \square
 \end{aligned}$$

**7.95. Theorem: Logarithm of a Product.** For all  $G, H \in K[[x]]$  with  $G(0) = 1 = H(0)$ ,

$$\log(GH) = \log(G) + \log(H).$$

More generally, given a sequence  $G_k \in K[[x]]$  with  $G_k(0) = 1$  for all  $k$  and  $\lim_{k \rightarrow \infty} G_k = 1$ ,

$$\log\left(\prod_{k=1}^{\infty} G_k\right) = \sum_{k=1}^{\infty} \log(G_k).$$

*Proof.* Since  $G$  and  $H$  have constant term 1, we know that

$$GH = \exp(\log G) \exp(\log H) = \exp[(\log G) + (\log H)].$$

Since  $GH$  has constant term 1, we can take logarithms to conclude that

$$\log(GH) = \log(G) + \log(H),$$

as desired. The formula for converting infinite products to infinite sums is proved similarly.  $\square$

Formal exponentials and logarithms obey formal differentiation rules entirely analogous to those learned in calculus.

**7.96. Theorem: Derivative Rules for Exponentials and Logarithms.** If  $G \in K[[x]]$  satisfies  $G(0) = 0$ , then  $(\exp(G))' = G' \exp G$ . If  $H \in K[[x]]$  satisfies  $H(0) = 1$ , then  $(\log(H))' = H'/H$ .

*Proof.* A direct calculation using the definition shows that  $E' = E + 1$ . Applying the formal chain rule, we conclude that

$$\begin{aligned} (\exp(G))' &= [(E \bullet G) + 1]' = (E' \bullet G)G' \\ &= ((E + 1) \bullet G)G' = ((E \bullet G) + 1)G' = G' \exp(G). \end{aligned}$$

Use this result to differentiate the identity  $\exp(\log(H)) = H$ . We obtain

$$(\log(H))' \exp(\log(H)) = H',$$

or equivalently  $(\log(H))' H = H'$ . Since  $H(0) = 1$ , we can divide by  $H$  to conclude that  $(\log(H))' = H'/H$ .  $\square$

**7.97. Theorem: Power Rule for Logarithms.** If  $H \in K[[x]]$  satisfies  $H(0) = 1$  and  $r \in K$ , then  $\log(H^r) = r \log(H)$ .

*Proof.* On one hand, both  $\log(H^r)$  and  $r \log(H)$  have formal derivative equal to  $rH'/H$  (by 7.79, 7.75, 7.96, and the chain rule). On the other hand, both  $\log(H^r)$  and  $r \log(H)$  have zero constant term. Thus these two series must be equal (see 7.132).  $\square$

## 7.16 Multivariable Polynomials and Formal Series

So far we have discussed polynomials and formal power series involving a single indeterminate. One can generalize this setup to arrive at the notions of multivariable polynomials and series.

**7.98. Definition: Formal Multivariable Power Series and Polynomials.** A *formal power series in  $k$  variables with coefficients in  $K$*  is a function  $F : \mathbb{N}^k \rightarrow K$ . The set of all such series is denoted  $K[[x_1, x_2, \dots, x_k]]$ . A series  $f \in K[[x_1, \dots, x_k]]$  is a *polynomial* iff  $\{\vec{n} \in \mathbb{N}^k : f(\vec{n}) \neq 0\}$  is finite. The set of all such polynomials is denoted  $K[x_1, \dots, x_k]$ .

The power series notation for a function  $F : \mathbb{N}^k \rightarrow K$  is

$$F = \sum_{(n_1, \dots, n_k) \in \mathbb{N}^k} F(n_1, \dots, n_k) x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k};$$

the function value  $F(n_1, \dots, n_k)$  is called the *coefficient of  $x_1^{n_1} \cdots x_k^{n_k}$  in  $F$* .  $F$  is a polynomial iff only a finite number of its coefficients are nonzero.

**7.99. Definition: Algebraic Operations on Multivariable Series.** Given  $c \in K$  and series  $F, G \in K[[x_1, \dots, x_k]]$ , the *sum*  $F + G$  is defined by  $(F + G)(\vec{n}) = F(\vec{n}) + G(\vec{n})$  for all  $\vec{n} \in \mathbb{N}^k$ . The *scalar multiple*  $cF$  is defined by  $(cF)(\vec{n}) = c(F(\vec{n}))$  for all  $\vec{n} \in \mathbb{N}^k$ . The *product*  $FG$  is defined by

$$(FG)(\vec{n}) = \sum_{\substack{(\vec{i}, \vec{j}) \in (\mathbb{N}^k)^2 \\ \vec{i} + \vec{j} = \vec{n}}} F(\vec{i})G(\vec{j}) \quad (\vec{n} \in \mathbb{N}^k).$$

**7.100. Example.** For  $1 \leq i \leq k$ , let  $x_i : \mathbb{N}^k \rightarrow K$  be the polynomial defined by sending  $(0, 0, \dots, 1, \dots, 0)$  (the 1 occurs in position  $i$ ) to 1 and sending everything else to zero. One can check that  $cx_1^{n_1} \cdots x_k^{n_k}$  is the series that sends  $(n_1, \dots, n_k)$  to  $c$  and everything else to zero. This justifies the notation used above for elements  $F \in K[[x_1, \dots, x_k]]$ , at least when  $F$  is a polynomial.

The following theorem is proved by making the necessary adjustments to the proofs given in the one-variable case. Even more general results are sketched in the exercises.

**7.101. Theorem: Algebraic Structure of Multivariable Series and Polynomials.**  $K[[x_1, \dots, x_k]]$  and  $K[x_1, \dots, x_k]$  are commutative rings that are integral domains, as well as vector spaces over  $K$  containing a copy of  $K$ . The set  $\{x_1^{n_1} \cdots x_k^{n_k} : (n_1, \dots, n_k) \in \mathbb{N}^k\}$  is a basis for the vector space  $K[x_1, \dots, x_k]$ .

Multivariable polynomial rings satisfy the following *universal mapping property* that generalizes 7.22. The proof is also left as an exercise.

**7.102. Theorem: Evaluation Homomorphisms for Multivariable Polynomials.**

Let  $S$  be a commutative ring containing  $K$ . (a) For each  $f \in K[x_1, \dots, x_k]$ , there is an associated function  $P_f : S^k \rightarrow S$  given by

$$P_f(z_1, \dots, z_k) = \sum_{(n_1, \dots, n_k) \in \mathbb{N}^k} f(n_1, \dots, n_k) z_1^{n_1} \cdots z_k^{n_k} \quad (z_i \in S).$$

(b) For each  $k$ -tuple  $\vec{z} = (z_1, \dots, z_k) \in S^k$ , there exists a unique ring homomorphism  $E : K[x_1, \dots, x_k] \rightarrow S$  such that  $E(c) = c$  for all  $c \in K$  and  $E(x_i) = z_i$  for all  $i \leq k$ ; namely,  $E(f) = P_f(\vec{z})$  for  $f \in K[x_1, \dots, x_k]$ . We write  $E = \text{ev}_{\vec{z}}$  and call it the *evaluation homomorphism determined by setting each  $x_i$  equal to  $z_i$* .

**7.103. Definition: Formal Partial Derivatives.** For  $1 \leq i \leq k$ , define a map  $D_i : K[[x_1, \dots, x_k]] \rightarrow K[[x_1, \dots, x_k]]$  by

$$D_i(F) = \frac{\partial F}{\partial x_i} = \sum_{(n_1, n_2, \dots, n_k) \in \mathbb{N}^k} (n_i + 1) F(n_1, \dots, n_i + 1, \dots, n_k) x_1^{n_1} \cdots x_i^{n_i} \cdots x_k^{n_k}.$$

$D_i$  is called the *formal partial derivative operator with respect to  $x_i$* .

It is routine to check that the analogues of the one-variable differentiation rules (§7.8) extend to the partial derivative operators  $D_i$ . There are also formal versions of the multi-variable chain rule. We now prove one such rule for multivariable polynomials, which will be used in §10.16.

**7.104. Theorem: Multivariable Chain Rule.** Let  $h \in K[y_1, \dots, y_n]$  and  $g_1, \dots, g_n \in K[x_1, \dots, x_m]$ . Let  $h \bullet g \in K[x_1, \dots, x_m]$  denote the polynomial obtained by setting each  $y_i = g_i$  in  $h$ . For  $1 \leq k \leq m$ ,

$$D_k(h \bullet g) = \sum_{j=1}^n ((D_j h) \bullet g) D_k(g_j).$$

Informally, we may write

$$\frac{\partial(h \bullet g)}{\partial x_k} = \frac{\partial h}{\partial y_1} \cdot \frac{\partial g_1}{\partial x_k} + \cdots + \frac{\partial h}{\partial y_n} \cdot \frac{\partial g_n}{\partial x_k},$$

with all of the partial derivatives  $\partial h / \partial y_j$  being evaluated at  $(y_1, \dots, y_n) = (g_1, \dots, g_n)$ .

*Proof.* Both sides of the claimed identity are  $K$ -linear functions of  $h$ . So it suffices to check the identity when  $h$  has the form  $y_1^{e_1} \cdots y_n^{e_n}$ . In this case,  $h \bullet g = g_1^{e_1} \cdots g_n^{e_n}$ . Viewing this as a product of  $e_1 + \cdots + e_n$  factors, each equal to some  $g_j$ , the multivariable product rule leads to

$$D_k(h \bullet g) = \sum_{j=1}^n e_j g_1^{e_1} \cdots g_j^{e_j-1} (D_k(g_j)) \cdots g_n^{e_n}.$$

On the other side,  $(D_j h) \bullet g = e_j g_1^{e_1} \cdots g_j^{e_j-1} \cdots g_n^{e_n}$ . Multiplying by  $D_k(g_j)$  and summing over  $j$  gives the same answer as before, so the proof is complete.  $\square$

## Summary

Table 7.1 reviews the definitions of concepts and operations involving formal power series and polynomials. Table 7.2 lists some rules and formulas arising in computations with formal power series (some hypotheses on constant terms are omitted in this table). Let us also recall the following results.

- *Algebraic Structure of  $K[[x]]$  and  $K[x]$ .* Both  $K[[x]]$  and  $K[x]$  are commutative rings, integral domains, and vector spaces over  $K$  containing a copy of  $K$ . The same holds for  $K[[x_1, \dots, x_k]]$  and  $K[x_1, \dots, x_k]$ . The set  $\{x^i : i \geq 0\}$  is a basis for  $K[x]$ , whereas the set  $\{x_1^{n_1} \cdots x_k^{n_k} : (n_1, \dots, n_k) \in \mathbb{N}^k\}$  is a basis for  $K[x_1, \dots, x_k]$ .
- *Degree and Order.* For polynomials  $f, g \in K[x]$ ,  $\deg(f + g) \leq \max(\deg(f), \deg(g))$  and  $\deg(fg) = \deg(f) + \deg(g)$  whenever both sides are defined. For series  $F, G \in K[[x]]$ , we have  $\text{ord}(F + G) \geq \min(\text{ord}(F), \text{ord}(G))$  and  $\text{ord}(FG) = \text{ord}(F) + \text{ord}(G)$  whenever both sides are defined.
- *Multiplicative Inverses in  $K[x]$  and  $K[[x]]$ .* A polynomial  $f$  is invertible (a unit) in  $K[x]$  iff  $\deg(f) = 0$ . A series  $F$  is invertible in  $K[[x]]$  iff  $F(0) \neq 0$ . In this case, one can use the formal geometric series to invert  $F$  or the recursive formula  $(F^{-1})_n = -(1/F_0) \sum_{k=1}^n F_k (F^{-1})_{n-k}$ . A nonzero formal quotient  $F/G \in K((x))$  can be written in a unique way as a Laurent series  $\sum_{n=m}^{\infty} a_n x^n$  where  $m \in \mathbb{Z}$ , each  $a_n \in K$ , and  $a_m \neq 0$ ;  $m = \text{ord}(F) - \text{ord}(G)$  is the *order* of this Laurent series.



**TABLE 7.1**

Definitions concerning formal power series and polynomials.

Term	Brief Definition
formal power series	function $F : \mathbb{N} \rightarrow K$ , denoted $\sum_{n=0}^{\infty} F_n x^n$
formal polynomial	formal series $f$ with $\{n : f_n \neq 0\}$ finite
ring	set with addition and multiplication satisfying axioms in 2.2
integral domain	nonzero commutative ring with no zero divisors
field	nonzero commutative ring with all nonzero elements invertible
$K[[x]]$	set of formal power series with coefficients in $K$
$K[x]$	set of formal polynomials with coefficients in $K$
$K(x)$	$\{f/g : f, g \in K[x], g \neq 0\}$ = field of fractions of $K[x]$
$K((x))$	$\{F/G : F, G \in K[[x]], G \neq 0\}$ = field of fractions of $K[[x]]$ , or the set of formal Laurent series $\sum_{n=m}^{\infty} a_n x^n$ ( $m \in \mathbb{Z}$ , $a_n \in K$ )
the series $x$	$x = X_1 = (0, 1, 0, 0, \dots) = \sum_{n=0}^{\infty} \chi(n=1) x^n$
equality of series	$F = G$ iff $F_n = G_n$ for all $n \in \mathbb{N}$
sum of series	$(F + G)(n) = F(n) + G(n)$ for all $n \in \mathbb{N}$
product of series	$(FG)(n) = \sum_{(i,j) \in \mathbb{N}^2: i+j=n} F(i)G(j)$ for $n \in \mathbb{N}$ (convolution)
derivative of series	$F'(n) = (dF/dx)_n = (n+1)F(n+1)$ for $n \in \mathbb{N}$
$k$ th derivative	$F^{(k)}(n) = (d^k F/dx^k)_n = (n+1) \cdots (n+k)F(n+k)$ ( $n \in \mathbb{N}$ )
formal limit	$F_n \rightarrow L$ iff $\forall m \in \mathbb{N}, \exists N \in \mathbb{N}, \forall n \in \mathbb{N}, (n \geq N \Rightarrow F_n(m) = L(m))$
infinite sum of series	$\sum_{n=0}^{\infty} F_n = \lim_{N \rightarrow \infty} \sum_{n=0}^N F_n$ (if limit exists)
infinite product	$\prod_{n=0}^{\infty} F_n = \lim_{N \rightarrow \infty} \prod_{n=0}^N F_n$ (if limit exists)
formal composition	$F \bullet G = \sum_{n=0}^{\infty} F_n G^n$ (need $F \in K[x]$ or $G(0) = 0$ )
formal $e^x$	$e^x = \sum_{n=0}^{\infty} x^n/n! \in K[[x]]$
formal $\sin x$	$\sin x = (0, 1, 0, -1/3!, 0, 1/5!, 0, \dots) \in K[[x]]$
formal $\cos x$	$\cos x = (1, 0, -1/2!, 0, 1/4!, 0, -1/6!, \dots) \in K[[x]]$
formal $\log(1+x)$	$\log(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} x^n/n \in K[[x]]$
formal $(1+x)^r$	$\text{Pow}_r = (1+x)^r = \sum_{n=0}^{\infty} ((r) \downarrow_n / n!) x^n \in K[[x]]$ ( $r \in K$ )
formal power	$F^r = \sum_{n=0}^{\infty} ((r) \downarrow_n / n!) (F-1)^n$ (need $F(0) = 1$ )
formal exponential	$\exp(G) = \sum_{n=0}^{\infty} G^n/n!$ (need $G(0) = 0$ )
formal logarithm	$\log(1+G) = \sum_{n=1}^{\infty} (-1)^{n-1} G^n/n$ (need $G(0) = 0$ ) or $\log(F) = \sum_{n=1}^{\infty} (-1)^{n-1} (F-1)^n/n$ (need $F(0) = 1$ )
degree	$\deg(f) = \max\{n : f_n \neq 0\}$ for nonzero $f \in K[x]$
order	$\text{ord}(F) = \min\{n : F_n \neq 0\}$ for nonzero $F \in K[[x]]$
polynomial function	for $f \in K[x]$ , $P_f : S \rightarrow S$ sends $z \in S \supseteq K$ to $\sum_{n=0}^{\deg(f)} f_n z^n$
ring homomorphism	map between rings preserving $+$ , $\times$ , and $1$
evaluation hom.	$\text{ev}_z(f) = P_f(z)$ for $f \in K[x]$ , $z \in S \supseteq K$
$K[[x_1, \dots, x_k]]$	set of $k$ -variable formal series $F : \mathbb{N}^k \rightarrow K$
$K[x_1, \dots, x_k]$	set of $k$ -variable polynomials
partial derivative	$D_i F(n_1, \dots, n_k) = (n_i + 1)F(n_1, \dots, n_i + 1, \dots, n_k)$

**TABLE 7.2**

Rules for calculating with formal power series.

Product of $k$ series:	$(F_1 \cdots F_k)(n) = \sum_{i_1 + \cdots + i_k = n} F_1(i_1) \cdots F_k(i_k).$
Positive powers:	$G^m(n) = \sum_{(k_0, \dots, k_n)}^m G(0)^{k_0} \cdots G(n)^{k_n},$ summed over $(k_0, \dots, k_n)$ with $\sum_i k_i = m$ and $\sum_i i k_i = n.$
Geometric series:	$(1 - G)^{-1} = \sum_{n=0}^{\infty} G^n$ when $G(0) = 0.$
Negative powers:	$(1 - G)^{-m} = \sum_{n=0}^{\infty} \binom{n+m-1}{n, m-1} G^n$ when $m \in \mathbb{N}^+$ and $G(0) = 0.$
Limit rules:	$F_n \rightarrow P$ and $G_n \rightarrow Q$ imply $F_n + G_n \rightarrow P + Q, \quad F_n G_n \rightarrow PQ, \quad F_n \bullet G_n \rightarrow P \bullet Q.$
Derivative rules:	$(F + G)' = F' + G', \quad (FG)' = (F')G + F(G'), \quad (x^k)' = kx^{k-1},$ $(F \bullet G)' = (F' \bullet G)G', \quad (F^r)' = rF^{r-1}F',$ $(\exp(G))' = G' \exp(G) \ [G(0) = 0], \quad (\log(H))' = H'/H \ [H(0) = 1],$ $(H_n \rightarrow F) \Rightarrow (H'_n \rightarrow F'), \quad (\sum_{n=1}^{\infty} H_n)' = \sum_{n=1}^{\infty} H'_n,$ $F = \sum_{k=0}^{\infty} (F^{(k)}(0)/k!)x^k; \quad D_k(h \bullet g) = \sum_j ((D_j h) \bullet g) D_k(g_j).$
Laws of exponents:	$\text{Pow}_{r+s} = \text{Pow}_r \text{Pow}_s; \quad F^{r+s} = F^r F^s;$ $(F^r)^n = F^{rn}$ for $r, s \in K, n \in \mathbb{Z}.$
Exp and Log:	$\exp(G + H) = \exp(G) \exp(H); \quad \exp(\sum_{k=1}^{\infty} G_k) = \prod_{k=1}^{\infty} \exp(G_k);$ $\log(GH) = \log(G) + \log(H); \quad \log(\prod_{k=1}^{\infty} G_k) = \sum_{k=1}^{\infty} \log(G_k);$ $\log(\exp(G)) = G \ [G(0) = 0]; \quad \exp(\log(H)) = H \ [H(0) = 1];$ $\log(H^r) = r \log(H) \ [H(0) = 1].$

- *Compositional Inverses in  $K[[x]]$ .* Formal composition is associative and has  $x$  as a two-sided identity. For fixed  $G$ , the map  $F \mapsto F \bullet G$  is a ring homomorphism fixing  $K$ . A series  $F \in K[[x]]$  has an inverse  $G$  relative to formal composition if  $F(0) = 0$  and  $F(1) \neq 0$ . The set of all such series is closed under composition and forms a group under this operation. The inverse  $G$  of  $F$  may be found by Lagrange inversion (see 8.15) or by the recursive formula  $G_n = -(1/F_1^n)(\sum_{m=0}^{n-1} G_m F^m)_n$ .
- *Evaluation Homomorphisms.* Let  $S$  be a commutative ring containing  $K$  and  $z_1, \dots, z_k \in S$ . There exists a unique ring homomorphism  $E : K[x_1, \dots, x_k] \rightarrow S$  such that  $E(x_i) = z_i$  for  $1 \leq i \leq k$  and  $E(c) = c$  for all  $c \in K$ .
- *Density of Polynomials in  $K[[x]]$ .* For each  $F \in K[[x]]$ ,  $F = \lim_{N \rightarrow \infty} \sum_{n=0}^N F_n x^n$ , so that any formal power series is a limit of formal polynomials.
- *Existence Criteria for Formal Limits.* For nonzero series  $F_k \in K[[x]]$ :  
 $F_k \rightarrow 0$  in  $K[[x]]$  iff  $\text{ord}(F_k) \rightarrow \infty$  in  $\mathbb{R}$ ;  
 $\sum_{k=0}^{\infty} F_k$  exists in  $K[[x]]$  iff  $\text{ord}(F_k) \rightarrow \infty$  in  $\mathbb{R}$ ;  
if  $F_k(0) = 0$  for all  $k$ ,  $\prod_{k=0}^{\infty} (1 + F_k)$  exists in  $K[[x]]$  iff  $\text{ord}(F_k) \rightarrow \infty$  in  $\mathbb{R}$ .
- *Polynomial Factorization in  $\mathbb{C}[x]$  and Partial Fractions.* A monic polynomial  $p \in \mathbb{C}[x]$  factors uniquely as  $(x - r_1)^{n_1} \cdots (x - r_k)^{n_k}$  with  $r_i \in \mathbb{C}$ . If instead  $p(0) = 1$ , we can write

$p = (1 - s_1x)^{n_1} \cdots (1 - s_kx)^{n_k}$  with  $s_i \in \mathbb{C}$ . Given  $q \in \mathbb{C}[x]$ , there is a unique expression

$$\frac{q}{p} = h + \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{a_{ij}}{(1 - s_ix)^j},$$

where  $h$  is the remainder when  $q$  is divided by  $p$ ; each  $a_{i,n_i}$  is found by multiplying all terms by  $(1 - s_ix)^{n_i}$  and setting  $x = 1/s_i$ ; and the remaining  $a_{ij}$ 's are found by subtracting the previously recovered terms and iterating.

- *Recursions with Constant Coefficients.* Suppose  $F_n = \sum_{i=1}^k c_i F_{n-i} + H_n$  for all  $n \geq k$ , where  $c_1, \dots, c_k \in K$  and  $H \in K[[x]]$  are given.  $F$  is uniquely determined by the  $k$  initial values  $F_0, \dots, F_{k-1}$ . The series  $F$  has the form  $G/p$ , where  $p = 1 - c_1x - \cdots - c_kx^k$  and  $G$  is a series with  $G_n = H_n$  for  $n \geq k$ .

## Exercises

**7.105.** Let  $f = x - x^2 + 3x^4$  and  $g = 1 - 2x + 3x^4$ . Compute  $f + g$ ,  $fg$ , and the degrees and orders of  $f$ ,  $g$ ,  $f + g$ , and  $fg$ .

**7.106.** Let  $F = (1, 0, 1, 0, 1, 0, \dots)$  and  $G = \sum_{n \geq 0} nx^n$ . Compute  $F + G$ ,  $FG$ ,  $F(1 + x)$ ,  $F(1 - x^2)$ ,  $G(1 + x)$ ,  $F'$ ,  $G'$ , and the orders of all these series.

**7.107.** Let  $f = x^2 + 4x - 1$  and  $g = x^3 + x$ . Compute  $P_f(2)$ ,  $P_g(\sqrt{5})$ ,  $P_f(x)$ ,  $P_f(g)$ ,  $P_g(f)$ , and  $P_f(f)$ .

**7.108.** Compute the coefficient of  $x^n$  for  $0 \leq n \leq 6$  for each of the following formal series: (a)  $e^x + \sin x$ ; (b)  $e^x \sin x$ ; (c)  $(\cos x) \log(1 + x)$ ; (d)  $(\log(1 + x))^2$ .

**7.109.** (a) Find necessary and sufficient conditions for strict inequality to hold in the formula  $\deg(f + g) \leq \max(\deg(f), \deg(g))$ . (b) Find necessary and sufficient conditions for strict inequality to hold in the formula  $\text{ord}(F + G) \geq \min(\text{ord}(F), \text{ord}(G))$ .

**7.110.** Use (7.3) in the proof of 7.40 to find the first five terms in the multiplicative inverse of each of the following series: (a)  $e^x$ ; (b)  $1 - 2x + x^3 + 3x^4$ ; (c)  $1 + \log(1 + x)$ .

**7.111.** Use 7.41 to find the first five terms in  $(1 - x + x^3)^{-1}$ .

**7.112.** Compute the multiplicative inverse of  $\sum_{n=0}^{\infty} n^2 x^n$  in  $K((x))$ .

**7.113.** Convert the following expressions to formal Laurent series: (a)  $(x^2 + 3)/(x^3 - x^2)$ ; (b)  $x/(x^3 - 5x^2 + 6x)$ .

**7.114. Formal Hyperbolic Sine and Cosine Functions.** Define formal series  $\sinh x = (e^x - e^{-x})/2$  and  $\cosh x = (e^x + e^{-x})/2$ . (a) Find  $(\sinh x)_n$  and  $(\cosh x)_n$  for all  $n \in \mathbb{N}$ . (b) Show  $(\sinh x)' = \cosh x$  and  $(\cosh x)' = \sinh x$ .

**7.115.** Complete the proof of 7.8(a) by verifying the remaining ring axioms for  $K[[x]]$ . Indicate which of the ring axioms for  $K$  are used in each part of your proof.

**7.116.** Let  $R$  be any ring. Verify that the sum and product operations in 7.6 (with  $K$  replaced by  $R$ ) make  $R[x]$  and  $R[[x]]$  rings, which are commutative if  $R$  is commutative.

**7.117.** This exercise shows that the characterization of units in  $K[x]$  given in 7.38 can fail if  $K$  is not a field. (a) Give an example of a commutative ring  $R$  and  $f \in R[x]$  such that  $\deg(f) = 0$  but  $f$  is not a unit of  $R[x]$ . (b) Give an example of a commutative ring  $R$  and  $f \in R[x]$  such that  $\deg(f) > 0$ , yet  $f$  is a unit of  $R[x]$ . (c) Show that for any  $n \in \mathbb{N}^+$ , there exists  $f$  as in part (b) with  $n$  nonzero coefficients.

**7.118. Continuity of Exp and Log.** (a) Assume  $F_k(0) = G(0) = 0$  and  $F_k \rightarrow G$  in  $K[[x]]$ . Prove  $\exp(F_k) \rightarrow \exp(G)$ . (b) Assume  $F_k(0) = G(0) = 1$  and  $F_k \rightarrow G$ . Prove  $\log(F_k) \rightarrow \log(G)$ .

**7.119.** Prove the following general version of the *universal mapping property for polynomial rings*. Let  $f : L \rightarrow R$  be a given ring homomorphism between two commutative rings. Given  $(z_1, \dots, z_k) \in R^k$ , there exists a unique ring homomorphism  $E : L[x_1, \dots, x_k] \rightarrow R$  such that  $E(x_i) = z_i$  for all  $i$  and  $E(c) = f(c)$  for all  $c \in L$ . Point out any steps in your proof that require the assumption that the rings are commutative.

**7.120.** (a) Show that, because  $K$  is an infinite field, the map  $\pi : K[x] \rightarrow {}^K K$ , given by  $\pi(f) = P_f$  for  $f \in K[x]$ , is injective. (b) Give an example of a commutative ring  $R$  such that the map  $\pi : R[x] \rightarrow {}^R R$  is not injective.

**7.121.** Prove 7.20(a),(c).

**7.122.** Let  $F_k, G_k, P, Q \in K[[x]]$  satisfy  $F_k \rightarrow P$  and  $G_k \rightarrow Q$ . Prove  $F_k + G_k \rightarrow P + Q$ .

**7.123.** Prove 7.54(a),(b),(c),(g).

**7.124.** Complete the following outline to give a new proof of the formal product rule  $(FG)' = (F')G + F(G')$  for  $F, G \in K[[x]]$ . (a) Show that the result holds when  $F = x^i$  and  $G = x^j$ , for all  $i, j \in \mathbb{N}$ . (b) Deduce from (a) that the result holds for all  $F, G \in K[x]$ . (c) Use a continuity argument to obtain the result for all  $F, G \in K[[x]]$ .

**7.125.** Prove 7.62(b),(c).

**7.126.** Use a continuity argument to deduce the formal chain rule for formal power series (see 7.64) from the chain rule for polynomials.

**7.127.** Prove the following formal derivative identities: (a)  $(\sin x)' = \cos x$ ; (b)  $(\cos x)' = -\sin x$ ; (c)  $[\log(1+x)]' = (1+x)^{-1}$ .

**7.128. Formal Quotient Rule.** Suppose  $F, G \in K[[x]]$  where  $G(0) \neq 0$ . Prove the derivative rule  $(F/G)' = (GF' - FG')/G^2$ .

**7.129. Formal Integrals.** The *formal integral* or *antiderivative* of a series  $F \in K[[x]]$  is the series

$$\int F dx = \sum_{n \geq 1} \frac{F_{n-1}}{n} x^n \in K[[x]],$$

which has constant term zero. Compute the formal integrals of the following formal power series: (a)  $3 + 2x - 7x^2 + 12x^5$ ; (b)  $\sum_{n \geq 0} n^2 x^n$ ; (c)  $\sum_{n \geq 0} (n+1)! x^n$ ; (d)  $e^x$ ; (e)  $\sin x$ ; (f)  $\cos x$ ; (g)  $(1+x)^{-1}$ ; (h)  $\frac{3+2x}{1-3x+2x^2}$ .

**7.130.** Prove the following facts about formal integrals (as defined in 7.129).

(a) (sum rule)  $\int F + G dx = \int F dx + \int G dx$  for  $F, G \in K[[x]]$ .

(b) (scalar rule)  $\int cF dx = c \int F dx$  for  $c \in K$  and  $F \in K[[x]]$ .

(c) (linear combination rule)  $\int \sum_{i=1}^n c_i H_i dx = \sum_{i=1}^n c_i \int H_i dx$  for  $c_i \in K$  and  $H_i \in K[[x]]$ . Can you formulate a similar statement for infinite sums?

- (d) (power rule)  $\int x^k dx = \frac{1}{k+1}x^{k+1}$  for all  $k \geq 0$ .  
 (e) (general antiderivatives) For all  $F, G \in K[[x]]$ ,  $G' = F$  iff there exists  $c \in K$  with  $G = \int F dx + c$ .  
 (f) (formal fundamental theorems of calculus)  $F = \frac{d}{dx} \int F dx$  and  $\int F' dx = F - F(0)$  for  $F \in K[[x]]$ .  
 (g) (continuity of integration) If  $F_k, H \in K[[x]]$  and  $F_k \rightarrow H$ , then  $\int F_k dx \rightarrow \int H dx$ .

**7.131.** Formulate and prove an “integration by parts” rule and a “substitution rule” for formal integrals (as defined in 7.129).

**7.132.** (a) Given  $F, G \in K[[x]]$ , prove that  $F = G$  iff  $F' = G'$  and  $F(0) = G(0)$ . (b) State and prove an analogous statement for multivariable series.

**7.133.** (a) Prove that  $(\sin x)^2 + (\cos x)^2 = 1$  in  $K[[x]]$  by computing the coefficient of  $x^n$  on each side. (b) Prove that  $(\sin x)^2 + (\cos x)^2 = 1$  in  $K[[x]]$  by invoking 7.132 and derivative rules.

**7.134.** Prove that  $(\cosh x)^2 - (\sinh x)^2 = 1$  in  $K[[x]]$ .

**7.135. Cancellation in an Integral Domain.** Let  $R$  be a nonzero commutative ring. Prove that  $R$  is an integral domain iff the following *cancellation axiom* holds: for all  $a, b, c \in R$ ,  $ab = ac$  and  $a \neq 0$  imply  $b = c$ .

**7.136. Product Rule for Multiple Factors.** Let  $F_1, \dots, F_k \in K[[x]]$ . Prove that

$$\frac{d}{dx}(F_1 F_2 \cdots F_k) = \sum_{j=1}^k F_1 \cdots F_{j-1} \left( \frac{d}{dx} F_j \right) F_{j+1} \cdots F_k.$$

Does a version of this rule hold for infinite products?

**7.137.** Use 7.11 to prove the differentiation rule  $\frac{d}{dx}(G^m) = mG^{m-1}G'$  for  $G \in K[[x]]$  and  $m \in \mathbb{N}^+$  without using the formal chain rule.

**7.138.** For  $m, n \in \mathbb{N}^+$ , evaluate the sum

$$\sum_{\substack{(k_0, k_1, \dots, k_n) \in \mathbb{N}^{n+1}, \\ \sum_i k_i = m, \sum_i i k_i = n}} \frac{m!}{k_0! 0!^{k_0} k_1! 1!^{k_1} \cdots k_n! n!^{k_n}}.$$

**7.139.** Let  $F = \prod_{k=1}^{\infty} (1 - x^k)$ . Find  $F_n$  for  $0 \leq n \leq 22$ . Can you see a pattern?

**7.140.** Carefully justify the following calculation:

$$\prod_{n=1}^{\infty} (1 - x^{2n-1})^{-1} = \prod_{i=1}^{\infty} (1 - x^{2^i}) \prod_{j=1}^{\infty} (1 - x^j)^{-1} = \prod_{k=1}^{\infty} (1 + x^k).$$

In particular, explain why all the infinite products appearing here exist.

**7.141.** Find a necessary and sufficient condition on series  $F_k \in K[[x]]$  so that the infinite product  $\prod_{k=1}^{\infty} (1 + F_k)^{-1}$  exists.

**7.142.** Evaluate  $\prod_{n=0}^{\infty} (1 + x^{2^n})$ .

**7.143.** Verify the partial fraction expansion of  $F$  given in 7.91.

**7.144.** Write out the formal series for each of the following expressions: (a)  $(1-x)^{-5}$ ; (b)  $\sqrt{1+x}$ ; (c)  $1/\sqrt{1-x^2}$ ; (d)  $\sqrt[3]{1+3x}$ .

**7.145.** Compute the first four nonzero terms in the following series: (a)  $\sqrt{1+x+3x^2}$ ; (b)  $\sqrt[3]{\cos x}$ ; (c)  $(\sum_{n \geq 0} (n+1)^2 x^n)^{-5/2}$ .

**7.146.** Compute the first four nonzero terms in: (a)  $\exp(\sin x)$ ; (b)  $\log(\cos x)$ .

**7.147.** Find the partial fraction decomposition of  $F = (10+2x)/(1-2x-8x^2)$ , and use this to determine  $F(n)$  for all  $n$ .

**7.148.** Find the partial fraction decomposition of  $F = (1-7x)/(15x^2-8x+1)$ , and use this to determine  $F(n)$  for all  $n$ .

**7.149.** Find the partial fraction decomposition of  $F = (2x^3-4x^2-x-3)/(2x^2-4x+2)$ , and use this to determine  $F(n)$  for all  $n$ .

**7.150.** Find the partial fraction decomposition of  $F = (15x^6+30x^5-15x^4-35x^4-15x^2-12x-8)/(15(x^4+2x^3-2x-1))$ , and use this to determine  $F(n)$  for all  $n$ .

**7.151.** (a) Solve the recursion  $a_n = 3a_{n-1}$  ( $n \geq 1$ ), given that  $a_0 = 2$ . (b) Solve the recursion  $a_n = 3a_{n-1} + 3n$  ( $n \geq 1$ ), given that  $a_0 = 2$ . (c) Solve the recursion  $a_n = 3a_{n-1} + 3^n$  ( $n \geq 1$ ), given that  $a_0 = 2$ .

**7.152.** Solve the recursion  $a_n = 6a_{n-1} - 8a_{n-2} + g(n)$  (for  $n \geq 2$ ) with initial conditions  $a_0 = 0$ ,  $a_1 = 2$  for the following choices of  $g(n)$ : (a)  $g(n) = 0$ ; (b)  $g(n) = 1$ ; (c)  $g(n) = 2^n$ ; (d)  $g(n) = n4^n$ .

**7.153.** The *Lucas numbers* are defined by setting  $L_0 = 1$ ,  $L_1 = 3$ , and  $L_n = L_{n-1} + L_{n-2}$  for  $n \geq 2$ . Use formal series to find a closed formula for  $L_n$ .

**7.154.** Solve the recursion  $a_n = -3a_{n-1} + 2a_{n-2} + 6a_{n-3} - a_{n-4} - 3a_{n-5}$  (for  $n \geq 5$ ) with initial conditions  $a_k = k$  for  $0 \leq k < 5$ .

**7.155.** Repeat 7.154 with initial conditions  $a_k = 3$  for  $0 \leq k < 5$ .

**7.156.** Suppose  $b_0 = 1$  and  $b_n = b_0 + b_1 + \cdots + b_{n-1} + 1$  for all  $n \geq 1$ . Find  $\sum_{n \geq 0} b_n x^n$ .

**7.157.** Suppose  $(c_n : n \in \mathbb{Z})$  satisfies  $c_0 = 0$ ,  $c_1 = 1$ , and  $c_n = (c_{n-1} + c_{n+1})/L$  for all  $n \in \mathbb{Z}$ , where  $L \in \mathbb{R}^+$  is a constant. Find an explicit formula for  $c_n$ .

**7.158. Differentiation of Laurent Series.** Define a version of the formal derivative operator for the ring  $K((x))$  of formal Laurent series. Extend the derivative rules (in particular, the quotient rule) to this ring.

**7.159. Formal Tangent and Secant Functions.** Define formal series  $\sec x = 1/\cos x$  and  $\tan x = \sin x/\cos x$ . (a) Compute  $(\sec x)_n$  and  $(\tan x)_n$  for  $0 \leq n \leq 9$ . (A combinatorial interpretation of these coefficients is described in §12.8.) (b) Show that  $(\tan x)^2 + 1 = (\sec x)^2$ . (c) Show that  $(\tan x)' = (\sec x)^2$  and  $(\sec x)' = \tan x \sec x$ . (d) Show that  $(\tan x)_n = 0$  for all even  $n$  and  $(\sec x)_n = 0$  for all odd  $n$ . (e) Can you give similar definitions and results for  $\cot x$  and  $\csc x$ ?

**7.160. Substitution of  $rx$  for  $x$ .** Given  $F \in K[[x]]$  and nonzero  $r \in K$ , define  $F(rx)$  to be the formal composition  $F \bullet (rx)$ . Prove: (a)  $\sin(2x) = 2 \sin x \cos x$ ; (b)  $\cos(2x) = (\cos x)^2 - (\sin x)^2$ ; (c)  $\exp(rx) = \exp(x)^r$  for nonzero  $r \in K$ ; (d)  $\exp(ix) = \cos x + i \sin x$ ,  $\cos x = (\exp(ix) + \exp(-ix))/2$ , and  $\sin x = (\exp(ix) - \exp(-ix))/2i$  (assuming  $i = \sqrt{-1} \in K$ ).

**7.161. Even and Odd Formal Series.** A series  $F \in K[[x]]$  is *even* iff  $F(-x) = F$  (see 7.160);  $F$  is *odd* iff  $F(-x) = -F$ . (a) Show that  $F$  is even iff  $F_n = 0$  for all odd  $n$ , and  $F$  is odd iff  $F_n = 0$  for all even  $n$ . (b) Which formal trigonometric and hyperbolic trigonometric series are odd? Which are even? (c) Give rules for determining the parity (even or odd) of  $F + G$ ,  $FG$ , and (when defined)  $F^{-1}$ , given the parity of  $F$  and  $G$ .

**7.162.** Let  $R$  be a commutative ring. Recall that  $x \in R$  is a *unit* of  $R$  iff there exists  $y \in R$  with  $xy = yx = 1_R$ ;  $x$  is *nilpotent* iff there exists  $n \in \mathbb{N}^+$  with  $x^n = 0_R$ . (a) Suppose  $x \in R$  is arbitrary and  $z \in R$  is nilpotent. Prove  $xz$  is nilpotent. (b) Suppose  $x \in R$  is a unit of  $R$  and  $y \in R$  is nilpotent. Prove  $x + y$  is a unit of  $R$ . (c) Suppose  $x, y \in R$  are both nilpotent. Prove  $x + y$  is nilpotent. (d) Which results in (a), (b), and (c) hold if  $R$  is a non-commutative ring?

**7.163.** Let  $R$  be a nonzero commutative ring. Prove that  $f \in R[x]$  is a unit of  $R[x]$  iff  $f_0$  is a unit of  $R$  and  $f_n$  is nilpotent in  $R$  for all  $n > 0$ .

**7.164.** Use (7.6) in the proof of 7.65 to find the first several coefficients in the compositional inverses of each of the following series: (a)  $\sin x$ ; (b)  $\tan x$ ; (c)  $x/(1-x)$ .

**7.165.** Use the formal Maclaurin formula 7.55 to deduce the series expansions of  $e^x$ ,  $\sin x$ ,  $\cos x$ , and  $(1-rx)^{-1}$  starting from the rules for differentiating these formal series.

**7.166.** Taylor's formula states that (under suitable hypotheses on  $f : \mathbb{R} \rightarrow \mathbb{R}$ )  $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n$  for all  $x$  sufficiently close to  $a$ . Give two reasons why this formula is not meaningful (as written) for formal power series, when  $a \in K$  is nonzero.

**7.167.** (a) Show that  $\sum_{n \geq 0} x^{3n}/(3n)! = (1/3)e^x + (2/3)\cos(x\sqrt{3}/2)e^{-x/2}$  in  $\mathbb{C}[[x]]$ . (b) Try to find similar formulas for  $\sum_{n \geq 0} x^{3n+1}/(3n+1)!$  and  $\sum_{n \geq 0} x^{3n+2}/(3n+2)!$ .

**7.168. Exponential Generating Functions.** Given a sequence  $F = (F_n : n \geq 0) \in K[[x]]$ , the *exponential generating function* of this sequence is  $F^* = \sum_{n \geq 0} (F_n/n!)x^n \in K[[x]]$ . Prove that, for all  $F, G \in K[[x]]$ : (a)  $(F+G)^* = F^* + G^*$ ; (b)  $n!(F^*G^*)_n = \sum_{k=0}^n \binom{n}{k} F_k G_{n-k}$ ; (c)  $\frac{d}{dx}(F^*) = ((F - F(0))/x)^*$ .

**7.169. Sum of Squares via Formal Series.** The goal of this problem is to use series to derive a formula for  $\sum_{k=0}^n k^2$  without guessing the answer in advance. (a) Express the series  $\sum_{n \geq 0} n^2 x^n$  as a linear combination of the series  $(1-x)^{-1}$ ,  $(1-x)^{-2}$ , and  $(1-x)^{-3}$ . (b) Perform a suitable operation on the series in (a) to obtain an algebraic formula for the series  $\sum_{n \geq 0} (\sum_{k=0}^n k^2) x^n$ . (c) Extract the coefficient of  $x^n$  in (b) to obtain a formula for  $\sum_{k=0}^n k^2$  that is a polynomial of degree 3 in  $n$ . (d) Explain another way to solve this problem based on 7.90.

**7.170.** Use the method of 7.169 to evaluate the following sums for all  $n$ : (a)  $\sum_{k=0}^n k$ ; (b)  $\sum_{k=0}^n k^3$ ; (c)  $\sum_{k=0}^n 3^k$ .

**7.171.** Prove that for all  $k, n \in \mathbb{N}^+$ ,

$$1^k + 2^k + \cdots + n^k = \sum_{j=0}^k \frac{S(k+1, j+1)}{j+1} (-1)^{k+j} (n+1) \uparrow_{j+1}.$$

**7.172.** State and prove a version of the quadratic formula for solving  $AF^2 + BF + C = 0$ , where  $A, B, C \in K[[x]]$  are known series and  $F \in K[[x]]$  is unknown. What hypotheses must you impose on  $A, B, C$ ? Is the solution  $F$  unique?

**7.173. Recursion for Divide-and-Conquer Algorithms.** Many algorithms use a “divide-and-conquer” approach in which a problem of size  $n$  is divided into  $a$  subproblems of size  $n/b$ , and the solutions to these subproblems are then combined in time  $cn^k$  to give the solution to the original problem. Letting  $T(n)$  be the time needed to solve a problem of size  $n$ ,  $T(n)$  will satisfy the recursion  $T(n) = aT(n/b) + cn^k$  and initial condition  $T(1) = d$  (where  $a, b, c, d > 0$  and  $k \geq 0$  are given constants). Assume for simplicity that  $n$  ranges over powers of  $b$ . (a) Find a recursion and initial condition satisfied by  $S(m) = T(b^m)$ , where  $m$  ranges over  $\mathbb{N}$ . (b) Use formal series to solve the recursion in (a). Deduce that, for a suitable constant  $C$  and large enough  $n$ ,

$$T(n) \leq \begin{cases} Cn^k & \text{if } a < b^k \text{ (combining time dominates);} \\ Cn^k \log_2 n & \text{if } a = b^k \text{ (dividing and combining times balance);} \\ Cn^{\log_b a} & \text{if } a > b^k \text{ (time to solve subproblems dominates).} \end{cases}$$

**7.174. Merge Sort.** Suppose we wish to sort a given sequence of integers  $x_1, \dots, x_n$  into increasing order. Consider the following recursive method: if  $n = 1$ , the sequence is already sorted. For  $n > 1$ , divide the list into two halves, sort each half recursively, and merge the resulting sorted lists. Let  $T(n)$  be the time needed to sort  $n$  objects using this algorithm. Find a recursion satisfied by  $T(n)$ , and use 7.173 to show that  $T(n) \leq Cn \log_2 n$  for some constant  $C$ . (You may assume  $n$  ranges over powers of 2.)

**7.175. Fast Binary Multiplication.** (a) Given  $x = ak + b$  and  $y = ck + d$  (where  $a, b, c, d, k \in \mathbb{N}$ ), verify that  $xy = (ak + b)(ck + d) = ack^2 + bd + ((a + b)(c + d) - ac - bd)k$ . Take  $k = 2^n$  in this identity to show that one can multiply two  $2n$ -bit numbers by recursively computing three products of  $n$ -bit numbers and doing several binary additions. (b) Find a recursion describing the number of bit operations needed to multiply two  $n$ -bit numbers by the recursive method suggested in (a). (c) Solve the recursion in (b) to determine the time complexity of this recursive algorithm (you may assume  $n$  is a power of 2).

**7.176. Formal Linear Ordinary Differential Equations.** Suppose  $P, Q \in K[[x]]$  are given formal series, and we wish to find a formal series  $F \in K[[x]]$  satisfying the “linear ODE”  $F' + PF = Q$  and initial condition  $F(0) = c \in K$ . Solve this ODE by multiplying by the “integrating factor”  $\exp(\int P dx)$  and using the product rule to simplify the left side.

**7.177. Formal ODEs with Constant Coefficients.** For fixed  $c_1, \dots, c_k \in \mathbb{C}$ , let  $V$  be the set of all formal series  $F \in \mathbb{C}[[x]]$  satisfying the ODE

$$F^{(k)} + c_1 F^{(k-1)} + c_2 F^{(k-2)} + \dots + c_k F = 0. \quad (7.12)$$

The *characteristic polynomial* for this ODE is  $q = x^k + c_1 x^{k-1} + c_2 x^{k-2} + \dots + c_k \in \mathbb{C}[x]$ . Suppose  $q$  factors as  $(x - r_1)^{k_1} \dots (x - r_s)^{k_s}$  for certain  $k_i > 0$  and distinct  $r_i \in \mathbb{C}$ . (a) Show that the  $k$  series  $x^j \exp(r_i x)$  (for  $1 \leq i \leq s$  and  $0 \leq j < k_i$ ) lie in  $V$ . (b) Show that  $V$  is a complex vector space, and the  $k$  series in (a) form a basis for  $V$ . (c) Describe a procedure for expressing a given sequence  $F \in V$  as a linear combination of the sequences in the basis from part (a), given the “initial conditions”  $F(0), F'(0), \dots, F^{(k-1)}(0)$ . (d) Let  $W$  be the set of formal series  $G \in \mathbb{C}[[x]]$  satisfying the non-homogeneous ODE

$$G^{(k)} + c_1 G^{(k-1)} + c_2 G^{(k-2)} + \dots + c_k G = H,$$

where  $H \in \mathbb{C}[[x]]$  is a given series. If  $G^*$  is one particular series in  $W$ , show that  $W = \{F + G^* : F \in V\}$ .

**7.178. Characteristic Polynomial of a Recursion.** This problem sets up an analogy between recursions with constant coefficients and ordinary differential equations with



constant coefficients. For fixed  $c_1, \dots, c_k \in \mathbb{C}$ , let  $V$  be the set of all formal series  $(A_n : n \geq 0) \in \mathbb{C}[[x]]$  satisfying the recursion

$$A_n = c_1 A_{n-1} + c_2 A_{n-2} + \cdots + c_k A_{n-k} \quad (n \geq k). \quad (7.13)$$

The *characteristic polynomial* for this recursion is  $q = x^k - c_1 x^{k-1} - c_2 x^{k-2} - \cdots - c_k \in \mathbb{C}[x]$ . Suppose  $q$  factors as  $(x - r_1)^{k_1} \cdots (x - r_s)^{k_s}$  for certain  $k_i > 0$  and distinct  $r_i \in \mathbb{C}$ . (a) Show that the  $k$  sequences  $(n^j r_i^n : n \geq 0)$  (for  $1 \leq i \leq s$  and  $0 \leq j < k_i$ ) lie in  $V$ . (b) Show that  $V$  is a complex vector space, and the  $k$  sequences in (a) form a basis for  $V$ . (c) Describe a procedure for expressing a given sequence  $A \in V$  as a linear combination of the sequences in the basis from part (a). (Use the “initial conditions”  $A_0, \dots, A_{k-1}$ .) (d) Let  $W$  be the set of sequences  $(B_n : n \geq 0)$  satisfying

$$B_n = c_1 B_{n-1} + c_2 B_{n-2} + \cdots + c_k B_{n-k} + g(n) \quad (n \geq k),$$

where  $g(n)$  is a given function. If  $B^*$  is one particular sequence in  $W$ , show that  $W = \{A + B^* : A \in V\}$ .

**7.179.** Fill in the details of the construction of the field of fractions of an integral domain, which was sketched in 7.44. Specifically, show that: (a) the relation  $\sim$  on  $X$  is reflexive, symmetric, and transitive; (b) addition and multiplication on  $F$  are well defined; (c)  $F$ , with these operations, is a field; (d) the map  $i : D \rightarrow F$  is an injective ring homomorphism; (e)  $F$  satisfies the universal mapping property stated in 7.44.

**7.180. Localization.** The construction of fields of fractions can be generalized as follows. Let  $R$  be a commutative ring (not necessarily an integral domain), and let  $S \subseteq R$  be a subset such that  $1 \in S$  and  $xy \in S$  whenever  $x, y \in S$ . Our goal is to use  $R$  to construct a new ring in which every element of  $S$  becomes a unit.

Define an equivalence relation on  $X = R \times S$  by setting  $(a, s) \sim (b, t)$  iff there exists  $u \in S$  with  $u(at - bs) = 0$ . (a) Show that  $\sim$  is an equivalence relation on  $X$ ; let  $T$  be the set of equivalence classes. (b) Define addition and multiplication operations on  $T$ ; show that these operations are well defined and make  $T$  into a commutative ring. (c) Define  $i : R \rightarrow T$  by letting  $i(r)$  be the equivalence class of  $(r, 1)$  for  $r \in R$ . Show that  $i$  is a ring homomorphism (which may not be injective, however) such that  $i(s)$  is a unit in  $T$  for every  $s \in S$ . (d) Show  $T$  has the following universal property: if  $U$  is any commutative ring and  $j : R \rightarrow U$  any ring homomorphism such that  $j(s)$  is a unit in  $U$  for every  $s \in S$ , then there exists a unique ring homomorphism  $f : T \rightarrow U$  such that  $j = f \circ i$ .

## Notes

A detailed but rather technical treatment of the algebraic theory of polynomials and formal power series is given in Bourbaki [19, Ch. IV]. More facts concerning polynomials, integral domains, and related aspects of ring theory may be found in algebra texts [8, 70, 71]. Discussions of formal power series from a more combinatorial perspective may be found in Stanley [127, Ch. 1] and Wilf [139].

---

## The Combinatorics of Formal Power Series

---

Now that we have the technical machinery of formal power series at our disposal, we can resume our combinatorial agenda of studying infinite weighted sets. We will develop versions of the weighted sum and product rules in this setting. We will also explore the combinatorial significance of other operations on formal series, like composition and exponentiation. These techniques will be used to obtain deeper combinatorial information about objects studied earlier in the book, including trees, integer partitions, and set partitions.

---

### 8.1 Sum Rule for Infinite Weighted Sets

**8.1. Definition: Admissible Weighted Sets and Generating Functions.** Suppose  $S$  is a set with weight function  $\text{wt} : S \rightarrow \mathbb{N}$ . The weighted set  $(S, \text{wt})$  is called *admissible* iff for every  $n \geq 0$ , the set  $S_n = \{z \in S : \text{wt}(z) = n\}$  is finite. In this case, the *generating function* of the weighted set  $(S, \text{wt})$  is the formal power series

$$G_S = G_{S, \text{wt}} = \sum_{n=0}^{\infty} |S_n| x^n \in \mathbb{Q}[[x]].$$

Informally, this series represents  $\sum_{z \in S} x^{\text{wt}(z)}$ .

**8.2. Example.** Every *finite* weighted set  $S$  is admissible. Furthermore, the generating function for such a set is a *polynomial* in  $x$ , since  $|S_n| = 0$  for all large enough  $n$ . We studied generating functions of this type in Chapter 6.

**8.3. Example.** Let  $S$  be the set of all binary trees (with any number of vertices), and let  $\text{wt}(T)$  be the number of vertices in  $T$  for each tree  $T \in S$ . Then  $S_n$  consists of all binary trees with  $n$  vertices. Even without determining the precise cardinality of  $S_n$  (which we did in 2.36), one can check that  $S_n$  is finite for each  $n \geq 0$ . Thus,  $S$  is an admissible weighted set. In most applications, we will not have calculated the cardinality  $|S_n|$  in advance — the whole point of using generating functions is to help solve problems like this! But in most situations of interest, it will follow routinely from the nature of the objects and weights that  $S_n$  is finite for all  $n$ . So we will often omit explicit proofs that the weighted sets under consideration are indeed admissible.

**8.4. Example.** Let  $(S, \text{wt})$  be an admissible weighted set. Let  $T$  be any subset of  $S$  with the same weight function as  $S$ . Then  $T_n \subseteq S_n$  for all  $n$ , so that  $(T, \text{wt})$  is also admissible. Similarly, a finite disjoint union of admissible weighted sets is again admissible.

**8.5. Theorem: Weight-Preserving Bijection Rule.** Suppose  $(S, \text{wt}_1)$  and  $(T, \text{wt}_2)$  are two weighted sets such that there exists a *weight-preserving bijection*  $f : S \rightarrow T$  (i.e.,  $\text{wt}_2(f(s)) = \text{wt}_1(s)$  for all  $s \in S$ ). Then  $S$  is admissible iff  $T$  is admissible, and  $G_S = G_T$ .

*Proof.* Because  $f$  (and hence  $f^{-1}$ ) are weight-preserving,  $f$  restricts to bijections  $f_n : S_n \rightarrow T_n$  for each  $n \geq 0$ . So  $|S_n| = |T_n|$  for all  $n \geq 0$ , which implies the desired conclusions.  $\square$

At last we are ready for the most general version of the sum rule.

**8.6. Sum Rule for Infinite Weighted Sets.** Suppose  $(S, \text{wt})$  is an admissible weighted set that is the disjoint union of subsets  $\{T_i : i \in I\}$ , where the index set  $I$  is finite or equal to  $\mathbb{N}$ . Assume that for all  $i \in I$  and all  $x \in T_i$ ,  $\text{wt}_{T_i}(x) = \text{wt}_S(x)$ . Then

$$G_S = \sum_{i \in I} G_{T_i}.$$

*Proof.* Let us compute the coefficient of  $x^n$  on each side, for fixed  $n \geq 0$ . Write  $S_n$  (resp.  $(T_i)_n$ ) for the set of objects in  $S$  (resp.  $T_i$ ) of weight  $n$ . By assumption,  $S_n$  is a finite set which is the disjoint union of the (necessarily finite) sets  $(T_i)_n$ . Let  $I_n$  be the set of indices such that  $(T_i)_n$  is nonempty; then  $I_n$  must be finite, since  $S_n$  is finite. By the ordinary sum rule for finite sets,

$$|S_n| = \sum_{i \in I_n} |(T_i)_n|.$$

The left side is the coefficient of  $x^n$  in  $G_S$ , while the right side is evidently the coefficient of  $x^n$  in  $\sum_{i \in I} G_{T_i}$ , since the summands corresponding to  $i \notin I_n$  contribute zero to the coefficient of  $x^n$ . When  $I = \mathbb{N}$ , this argument also proves the convergence of the infinite sum of formal power series  $\sum_{i \in I} G_{T_i}$ , since the coefficient of  $x^n$  stabilizes once  $i \geq \max\{j : j \in I_n\}$ .  $\square$

## 8.2 Product Rule for Infinite Weighted Sets

In this section we prove two versions of the product rule for generating functions. The first version is designed for situations in which we build weighted objects by making a *finite* sequence of choices, as in Chapter 1. The second version extends this rule to certain infinite choice sequences, which leads to formulas involving infinite products of formal power series.

**8.7. Informal Product Rule for Infinite Weighted Sets.** Suppose  $(S, \text{wt})$  is a weighted set;  $k$  is a fixed, finite positive integer; and  $(T_i, \text{wt}_i)$  are admissible weighted sets for  $1 \leq i \leq k$ . Suppose each  $z \in S$  can be uniquely constructed by choosing  $z_1 \in T_1$ , then  $z_2 \in T_2$ , ..., then  $z_k \in T_k$ , and then assembling these choices in some manner. Further suppose that

$$\text{wt}(z) = \sum_{i=1}^k \text{wt}(z_i) \tag{8.1}$$

for all  $z \in S$ . Then  $(S, \text{wt})$  is admissible, and

$$G_S = \prod_{i=1}^k G_{T_i}.$$

*Proof.* Recast in formal terms, our hypothesis is that there is a weight-preserving bijection from  $(S, \text{wt})$  to the weighted set  $(T, \text{wt})$  where  $T = T_1 \times \cdots \times T_k$  and  $\text{wt}(z_1, \dots, z_k) = \sum_{i=1}^k \text{wt}(z_i)$ . So it suffices to replace  $S$  by the Cartesian product set  $T$ . Furthermore, it

suffices to prove the result when  $k = 2$ , since the general case follows by induction as in 1.5. Fix  $n \geq 0$ ; we are reduced to proving

$$G_{T_1 \times T_2}(n) = (G_{T_1} G_{T_2})(n).$$

The left side is the cardinality of the set  $A = \{(t_1, t_2) \in T_1 \times T_2 : \text{wt}_1(t_1) + \text{wt}_2(t_2) = n\}$ . Now  $A$  is the disjoint union of the sets  $(T_1)_k \times (T_2)_{n-k}$ , where  $(T_1)_k$  (resp.  $(T_2)_{n-k}$ ) is the finite set of objects in  $T_1$  (resp.  $T_2$ ) of weight  $k$  (resp.  $n - k$ ), and  $k$  ranges from 0 to  $n$ . So  $A$  is a finite set (proving admissibility of  $T_1 \times T_2$ ), and the ordinary sum and product rules for finite unweighted sets give

$$|A| = \sum_{k=0}^n |(T_1)_k| \cdot |(T_2)_{n-k}| = \sum_{k=0}^n G_{T_1}(k) \cdot G_{T_2}(n - k).$$

This sum is precisely the coefficient of  $x^n$  in  $G_{T_1} G_{T_2}$ , so we are done.  $\square$

The following technical device will allow us to obtain generating functions for objects that are built by making an *infinite* sequence of choices.

**8.8. Definition: Restricted Cartesian Product.** Suppose  $\{(T_n, \text{wt}_n) : n \geq 1\}$  is a countable collection of admissible weighted sets such that every  $T_n$  contains exactly one element of weight zero; call this element  $1_n$ . Let  $T = \prod_{n \geq 1}^* T_n$  be the set of all infinite sequences  $(t_n : n \geq 1)$  such that  $t_n \in T_n$  for all  $n$  and  $t_n = 1_n$  for all but finitely many indices  $n$ . We make  $T$  into a (not necessarily admissible) weighted set by defining  $\text{wt}((t_n : n \geq 1)) = \sum_{n \geq 1} \text{wt}(t_n)$ ; this sum is defined since all but finitely many summands are zero.

**8.9. Product Rule for the Restricted Cartesian Product.** Let  $\{(T_n, \text{wt}_n) : n \geq 1\}$  and  $T = \prod_{n \geq 1}^* T_n$  be as in 8.8. If  $\text{ord}(G_{T_n} - 1) \rightarrow \infty$  as  $n \rightarrow \infty$ , then  $(T, \text{wt})$  is admissible and

$$G_T = \prod_{n=1}^{\infty} G_{T_n}.$$

*Proof.* The condition on the orders of  $G_{T_n} - 1$  ensures that the infinite product of formal series is defined (see 7.33(c)). Fix  $m \geq 0$ ; let us compute  $G_T(m)$ . Choose  $N$  so that  $n > N$  implies  $\text{ord}(G_{T_n} - 1) > m$ . Consider an object  $t = (t_n : n \geq 1)$  in  $T$ . If  $t_n \neq 1_n$  for some  $n > N$ , then  $\text{wt}(t) \geq \text{wt}_n(t_n) > m$ , so this object does not contribute to the coefficient  $G_T(m)$ . So we need only consider objects where  $t_n = 1_n$  for all  $n > N$ . Dropping all coordinates after position  $N$  gives a weight-preserving bijection between this set of objects and the weighted set  $T_1 \times \cdots \times T_N$ . We already know that the generating function for this weighted set is  $\prod_{n=1}^N G_{T_n}$ . So

$$G_T(m) = \left( \prod_{n=1}^N G_{T_n} \right)_m = \left( \prod_{n=1}^{\infty} G_{T_n} \right)_m ;$$

the last equality holds since  $\text{ord}(G_{T_n} - 1) > m$  for  $n > N$ . This argument has shown that  $G_T(m)$  is finite for each  $m$ , which is equivalent to admissibility of  $T$ .  $\square$

To apply this result, we start with some weighted set  $(S, \text{wt})$  and describe an “infinite choice sequence” for building objects in  $S$  by choosing “building blocks” from the sets  $T_n$ . Each set  $T_n$  has a “dummy object” of weight zero. Any particular choice sequence must eventually terminate by choosing the dummy object for all sufficiently large  $n$ , but there is no fixed bound on the number of “non-dummy” choices we might make. This informal choice procedure amounts to giving a weight-preserving bijection from  $S$  to the restricted product  $\prod_{n \geq 1}^* T_n$ . We can then conclude that  $G_S = \prod_{n \geq 1} G_{T_n}$ , provided that the infinite product on the right side converges.

### 8.3 Generating Functions for Trees

This section illustrates the sum and product rules for infinite weighted sets by deriving the generating functions for various classes of trees.

**8.10. Example: Binary Trees.** Let  $S$  be the set of all binary trees, weighted by the number of vertices. By definition (see 2.36), every tree  $t \in S$  is either empty or is an ordered triple  $(\bullet, t_1, t_2)$ , where  $t_1$  and  $t_2$  are binary trees. Let  $S_0$  be the one-element set consisting of the empty binary tree, let  $S^+ = S \sim S_0$  be the set of nonempty binary trees, and let  $N = \{\bullet\}$  be a one-element set such that  $\text{wt}(\bullet) = 1$ . By definition of the generating function for a weighted set, we have  $G_{S_0} = x^0 = 1$  and  $G_N = x^1 = x$ . By the sum rule for infinite weighted sets,

$$G_S = G_{S_0} + G_{S^+} = 1 + G_{S^+}.$$

By the recursive definition of nonempty binary trees, we can uniquely construct every tree  $t \in S^+$  by: (i) choosing the root node  $\bullet \in N$ ; (ii) choosing the left subtree  $t_1 \in S$ ; (iii) choosing the right subtree  $t_2 \in S$ ; and assembling these choices to form the tree  $t = (\bullet, t_1, t_2)$ . It follows from the product rule for infinite weighted sets that

$$G_{S^+} = G_N G_S G_S = x G_S^2.$$

Writing  $F$  to denote the unknown generating function  $G_S$ , we conclude that  $F$  satisfies the equation

$$F = 1 + xF^2$$

in  $\mathbb{Q}[[x]]$ , which is equivalent to  $xF^2 - F + 1 = 0$ . Furthermore,  $F(0) = 1$  since there is exactly one binary tree with zero vertices. In 7.80, we saw that this quadratic equation and initial condition has the unique solution

$$F = \frac{1 - \sqrt{1 - 4x}}{2x} = \sum_{n=0}^{\infty} \frac{1}{2n+1} \binom{2n+1}{n+1, n} x^n.$$

Taking the coefficient of  $x^n$  gives the number of binary trees with  $n$  nodes, which is the Catalan number  $C_n$ . A more combinatorial approach to this result was given in Chapter 2.

**8.11. Example: Full Binary Trees.** A binary tree is called *full* iff every vertex in the tree has either zero or two (nonempty) children. In the context of binary trees, a *leaf* is a vertex with zero children. Let  $S$  be the set of nonempty full binary trees, weighted by the number of leaves. We can write  $S$  as the disjoint union of  $S_1 = \{(\bullet, \emptyset, \emptyset)\}$  and  $S_{\geq 2} = S \sim S_1$ . We can build an element  $t$  of  $S_{\geq 2}$  by choosing any  $t_1 \in S$  as the (nonempty) left subtree of the root, and then choosing any  $t_2 \in S$  as the right subtree of the root. Note that  $\text{wt}(t) = \text{wt}(t_1) + \text{wt}(t_2)$  since the weight is the number of leaves. So, by the product rule,  $G_{S_{\geq 2}} = G_S^2$ . We see directly that  $G_{S_1} = x$ . The sum rule now gives the relation

$$G_S = x + G_S^2,$$

with  $G_S(0) = 0$ . Solving the quadratic  $G_S^2 - G_S + x = 0$  by calculations analogous to those in 7.80, we find that

$$G_S = \frac{1 - \sqrt{1 - 4x}}{2} = xF,$$

where  $F$  is the generating function considered in the previous example. It follows that  $G_S(n) = F(n-1) = C_{n-1}$  for all  $n \geq 1$ .

**8.12. Example: Ordered Trees.** Let  $S$  be the set of *ordered trees*, weighted by the number of vertices. We recall the recursive definition of ordered trees from 3.79. First, 0 is an ordered tree with one vertex. Second, for every integer  $k \geq 1$ , a tuple  $t = (k, t_1, t_2, \dots, t_k)$  such that each  $t_i \in S$  is an ordered tree, and the number of vertices of  $t$  is  $1 + \sum_{i=1}^k \text{wt}(t_i)$ . (Informally,  $t$  represents a tree whose root has  $k$  children, which are ordered from left to right, and where each child is itself an ordered tree.) All ordered trees arise by applying the two rules a finite number of times. The first rule can be considered a degenerate version of the second rule in which  $k = 0$ . Let us find the generating function  $G_S$ . First, write  $S$  as the disjoint union of sets  $\{S_k : k \geq 0\}$  where  $S_k$  consists of all trees  $t \in S$  such that the root node has  $k$  children. By the sum rule for infinite weighted sets,

$$G_S = \sum_{k=0}^{\infty} G_{S_k}.$$

(One can verify the admissibility hypothesis on  $S$  by noting that every tree in  $S_k$  has  $k$  or more leaves.) For each  $k \geq 0$ , a direct application of the product rule (with  $k + 1$  choices) shows that

$$G_{S_k} = x^1 \cdot \underbrace{G_S \cdots G_S}_k = x G_S^k.$$

(The  $x$  arises by choosing the root node from the set  $\{\bullet\}$ .) Substitution into the previous formula gives

$$G_S = \sum_{k=0}^{\infty} x G_S^k = \frac{x}{1 - G_S};$$

the last step is valid (using 7.41) since  $G_S(0) = 0$ . Doing algebra in the ring  $\mathbb{Q}[[x]]$  leads to the relation  $G_S^2 - G_S + x = 0$ . This is the same equation that occurred in the previous example. So we conclude, as before, that

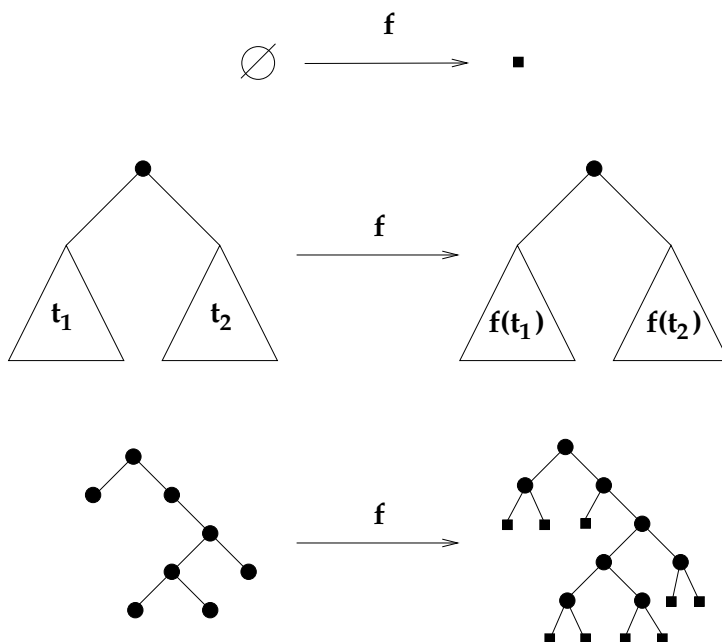
$$G_S = \frac{1 - \sqrt{1 - 4x}}{2} = \sum_{n \geq 1} C_{n-1} x^n.$$

Our generating function calculations have led us to the following (possibly unexpected) enumeration result: *the set of binary trees with  $n$  vertices, the set of full binary trees with  $n + 1$  leaves, and the set of ordered trees with  $n + 1$  vertices all have cardinality  $C_n$ .* Now that this result is in hand, it is natural to seek a *bijective proof* in which the three sets of objects are linked by explicitly defined bijections. Some methods for building such bijections from recursions were studied in Chapter 2. Here we are seeking weight-preserving bijections on infinite sets, which can be defined as follows.

Let  $S$  denote the set of all binary trees, and let  $T$  be the set of all nonempty full binary trees. We define a weight-preserving bijection  $f : S \rightarrow T$  recursively by setting  $f(\emptyset) = (\bullet, \emptyset, \emptyset)$  and

$$f((\bullet, t_1, t_2)) = (\bullet, f(t_1), f(t_2)).$$

See Figure 8.1. To see that the weights work, first note that the zero-vertex tree  $\emptyset$  is mapped to the one-leaf tree  $(\bullet, \emptyset, \emptyset)$ . In the recursive formula, suppose  $t_1$  and  $t_2$  have  $a$  vertices and  $b$  vertices, respectively. By induction,  $f(t_1)$  and  $f(t_2)$  are nonempty full binary trees with  $a + 1$  leaves and  $b + 1$  leaves, respectively. It follows that  $f$  sends the tree  $(\bullet, t_1, t_2)$  with  $a + b + 1$  vertices to a full binary tree with  $(a + 1) + (b + 1) = (a + b + 1) + 1$  leaves, as desired. The inverse of  $f$  has an especially simple pictorial description: just erase all the leaves! This works since a nonempty full binary tree always has one more leaf vertex than internal (non-leaf) vertex (see 8.50).

**FIGURE 8.1**

Bijection between binary trees and full binary trees.

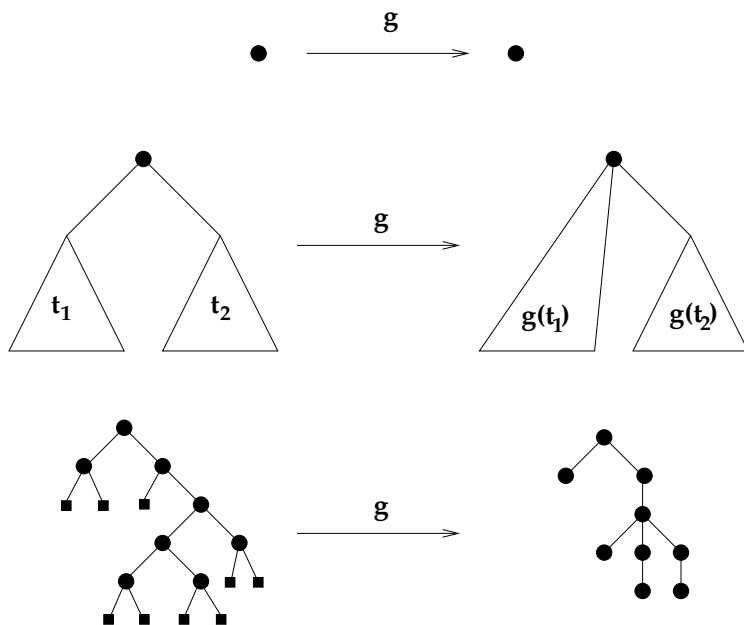
Now let  $U$  be the set of all ordered trees. We define a weight-preserving bijection  $g : T \rightarrow U$ . First,  $g((\bullet, \emptyset, \emptyset)) = 0$ . Second, if  $t = (\bullet, t_1, t_2) \in T$  with  $t_1$  and  $t_2$  nonempty, define  $g(t)$  by starting with the ordered tree  $g(t_1)$  and appending the ordered tree  $g(t_2)$  as a new, rightmost child of the root node of  $g(t_1)$ . See Figure 8.2. More formally, if  $g(t_1) = k u_1 \dots u_k$ , let  $g(t) = (k + 1) u_1 \dots u_k g(t_2)$ . As above, one may check that the number of vertices in  $g(t)$  equals the number of leaves in  $t$ , as required.

**8.13. Remark.** These examples show that generating functions are a powerful algebraic tool for deriving enumeration results. However, once such results are found, it is often desirable to find direct combinatorial proofs that do not rely on generating functions. In particular, bijective proofs are more informative (and often more elegant) than algebraic proofs in the sense that they give us an explicit pairing between the objects in two sets.

## 8.4 Compositional Inversion Formulas

Let  $F = \sum_{n \geq 1} F_n x^n$  be a formal series with  $F(0) = 0$  and  $F_1 \neq 0$ . We have seen in 7.65 that there is a unique series  $G = \sum_{n \geq 1} G_n x^n$  with  $G(0) = 0$  and  $G_1 \neq 0$  such that  $F \bullet G = x = G \bullet F$ . Our goal in this section is to find combinatorial and algebraic formulas for the coefficients of  $G$ .

Since  $F_1 \neq 0$ , we can write  $F = x/R$  where  $R = \sum_{n \geq 0} R_n x^n$  is a series with  $R_0 \neq 0$  (see 7.40). We have  $F \bullet G = x$  iff  $G/(R \bullet G) = x$  iff  $G = x(R \bullet G)$ . It turns out that we can solve the equation  $G = x(R \bullet G)$  by taking  $G$  to be the generating function for the set

**FIGURE 8.2**

Bijection between full binary trees and ordered trees.

of ordered trees (or equivalently, terms), relative to a suitable weight function. This is the essence of the following combinatorial formula for  $G$ .

**8.14. Theorem: Combinatorial Compositional Inversion.** Let  $F = x/R$  where  $R = \sum_{n \geq 0} R_n x^n$  is a given series in  $K[[x]]$  with  $R_0 \neq 0$ . Let  $T$  be the set of all terms (§3.13). Let the *weight* of a term  $w = w_1 w_2 \cdots w_s \in T$  be  $\text{wt}(w) = R_{w_1} R_{w_2} \cdots R_{w_s} x^s$ . Then  $G = G_T = \sum_{w \in T} \text{wt}(w)$  is the compositional inverse of  $F$ .

*Proof.* Note first that, for any two words  $v$  and  $w$ ,  $\text{wt}(vw) = \text{wt}(v) \text{wt}(w)$ . Also  $G(0) = 0$ , since every term has positive length. Now, by 3.85 we know that for every term  $w \in T$ , there exist a unique integer  $n \geq 0$  and unique terms  $t_1, \dots, t_n \in T$  such that  $w = n t_1 t_2 \cdots t_n$ . For fixed  $n$ , we build such a term by choosing the symbol  $n$  (which has weight  $x R_n$ ), then choosing terms  $t_1 \in T$ ,  $t_2 \in T$ ,  $\dots$ ,  $t_n \in T$ . By the product rule for generating functions (which generalizes to handle the current weights), the generating function for terms starting with  $n$  is therefore  $x R_n G^n$ . Now by the sum rule, we conclude that

$$G = \sum_{n \geq 0} x R_n G^n = x(R \bullet G).$$

By the remarks preceding the theorem, this shows that  $F \bullet G = x$ , as desired.  $\square$

In 3.91 we gave a formula that counts all terms in a given anagram class  $\mathcal{R}(0^{k_0} 1^{k_1} 2^{k_2} \cdots)$ . Combining this formula with the previous result, we deduce the following algebraic recipe for the coefficients of  $G$ .

**8.15. Theorem: Lagrange's Inversion Formula.** Let  $F = x/R$  where  $R = \sum_{n \geq 0} R_n x^n$  is a given series in  $K[[x]]$  with  $R_0 \neq 0$ . Let  $G$  be the compositional inverse of  $F$ . Then

$$G(n) = (R^n)_{n-1}/n = \frac{1}{n!} \left[ \left( \frac{d}{dx} \right)^{n-1} R^n \right]_0 \quad (n \geq 1).$$



*Proof.* The second equality follows routinely from the definition of formal differentiation. As for the first, let  $T_n$  be the set of terms of length  $n$ . By 8.14, we know that

$$G(n) = \sum_{w \in T_n} \text{wt}(w).$$

Let us group together summands on the right side corresponding to terms of length  $n$  that contain  $k_0$  zeroes,  $k_1$  ones, etc., where  $\sum_{i \geq 0} k_i = n$ . Each such term has weight  $x^n R_0^{k_0} R_1^{k_1} \cdots$ , and the number of such terms is  $\frac{1}{n} \binom{n}{k_0, k_1, k_2, \dots}$ , provided that  $k_0 = 1 + \sum_{i \geq 1} (i-1)k_i$  (3.91). Summing over all possible choices of the  $k_i$ , we get

$$G(n) = \sum_{\substack{(k_0, k_1, k_2, \dots): \\ \sum_{i \geq 0} k_i = n, \quad k_0 = 1 + \sum_{i \geq 1} (i-1)k_i}} \frac{1}{n} \binom{n}{k_0, k_1, k_2, \dots} R_0^{k_0} R_1^{k_1} R_2^{k_2} \cdots.$$

Now, in the presence of the condition  $\sum_{i \geq 0} k_i = n$ , the equation  $k_0 = 1 + \sum_{i \geq 1} (i-1)k_i$  holds iff  $\sum_{i \geq 0} (i-1)k_i = -1$  iff  $\sum_{i \geq 0} ik_i = n-1$ . So

$$G(n) = \sum_{\substack{(k_0, k_1, k_2, \dots): \\ \sum_{i \geq 0} k_i = n, \quad \sum_{i \geq 0} ik_i = n-1}} \frac{1}{n} \binom{n}{k_0, k_1, k_2, \dots} R_0^{k_0} R_1^{k_1} R_2^{k_2} \cdots.$$

On the other hand, 7.11 gives

$$(R_{n-1}^n)/n = \sum_{\substack{(k_0, k_1, k_2, \dots): \\ \sum_{i \geq 0} k_i = n, \quad \sum_{i \geq 0} ik_i = n-1}} \frac{1}{n} \binom{n}{k_0, k_1, k_2, \dots} R_0^{k_0} R_1^{k_1} R_2^{k_2} \cdots.$$

The right sides agree, so we are done.  $\square$

**8.16. Example.** Let us use 8.15 to find the compositional inverse  $G$  of the formal series  $F = x/e^x$ . Here  $R = e^x = \sum_{k \geq 0} x^k/k!$ , so  $R^n = e^{nx} = \sum_{k \geq 0} (n^k/k!)x^k$ . It follows that

$$G(n) = (R^n)_{n-1}/n = \frac{n^{n-1}}{n \cdot (n-1)!} = \frac{n^{n-1}}{n!}.$$

Thus,  $G = \sum_{n \geq 1} (n^{n-1}/n!)x^n$ .

## 8.5 Generating Functions for Partitions

This section uses formal power series to prove some fundamental results involving integer partitions. Recall from §2.8 that  $\text{Par}$  denotes the set of all integer partitions. Our first result gives an infinite product formula for the partition generating function.

### 8.17. Theorem: Partition Generating Function.

$$\sum_{\mu \in \text{Par}} x^{|\mu|} = \prod_{i=1}^{\infty} \frac{1}{1-x^i}.$$

*Proof.* The proof is an application of the infinite product rule 8.9. We build a typical partition  $\mu \in \text{Par}$  by making an infinite sequence of choices, as follows. First, choose how many parts of size 1 will occur in  $\mu$ . The possible choices here are  $0, 1, 2, 3, \dots$ . The generating function for this choice (relative to area) is  $1 + x + x^2 + x^3 + \dots = (1 - x)^{-1}$ . Second, choose how many parts of size 2 will occur in  $\mu$ . Again the possibilities are  $0, 1, 2, 3, \dots$ , and the generating function for this choice is  $1 + x^2 + x^4 + x^6 + \dots = (1 - x^2)^{-1}$ . Proceed similarly, choosing for every  $i \geq 1$  how many parts of size  $i$  will occur in  $\mu$ . The generating function for choice  $i$  is  $\sum_{k=0}^{\infty} (x^i)^k = (1 - x^i)^{-1}$ . Multiplying the generating functions for all the choices gives the infinite product in the theorem.

Here is a more formal rephrasing of the proof just given. For each  $i \geq 1$ , let  $T_i$  be the set of all integer partitions  $\nu$  (including the empty partition of zero) such that every part of  $\nu$  is equal to  $i$ . As argued above,  $G_{T_i} = (1 - x^i)^{-1}$ . Given any partition  $\mu$ , write  $\mu = (1^{a_1} 2^{a_2} \dots i^{a_i} \dots)$  to indicate that  $\mu$  has  $a_i$  parts equal to  $i$  for all  $i$ . (Note that  $a_i = 0$  for large enough  $i$ .) Then the map  $(1^{a_1} 2^{a_2} \dots i^{a_i} \dots) \mapsto ((1^{a_1}), (2^{a_2}), \dots, (i^{a_i}), \dots)$  is a weight-preserving bijection from  $\text{Par}$  onto  $\prod_{i \geq 1}^* T_i$ . The result now follows directly from 8.9.  $\square$

We can add another variable to the partition generating function to keep track of additional information. Recall that, for  $\mu \in \text{Par}$ ,  $\mu_1$  is the length of the first (longest) part of  $\mu$ , and  $\ell(\mu)$  is the number of nonzero parts of  $\mu$ .

### 8.18. Theorem: Enumerating Partitions by Area and Length.

$$\sum_{\mu \in \text{Par}} t^{\ell(\mu)} x^{|\mu|} = \prod_{i=1}^{\infty} \frac{1}{1 - tx^i} = \sum_{\mu \in \text{Par}} t^{\mu_1} x^{|\mu|} \quad \text{in } \mathbb{Q}(t)[[x]].$$

*Proof.* To prove the first equality, we modify the preceding argument to take into account the  $t$ -weight. At stage  $i$ , suppose we choose  $k$  copies of the part  $i$  for inclusion in  $\mu$ . This will increase  $\ell(\mu)$  by  $k$  and increase  $|\mu|$  by  $ki$ . So the generating function for the choice made at stage  $i$  is

$$\sum_{k \geq 0} t^k x^{ki} = \sum_{k \geq 0} (tx^i)^k = \frac{1}{1 - tx^i}.$$

The result now follows from the product rule, as before. To prove the second equality, observe that conjugation is a bijection on  $\text{Par}$  that preserves area and satisfies  $(\mu')_1 = \ell(\mu)$ .  $\square$

We can use variations of the preceding arguments to derive generating functions for various classes of integer partitions.

**8.19. Theorem: Partitions with Odd Parts.** Let  $\text{OddPar}$  be the set of integer partitions all of whose parts are odd. Then

$$\sum_{\mu \in \text{OddPar}} x^{|\mu|} = \prod_{k=1}^{\infty} \frac{1}{1 - x^{2k-1}}.$$

*Proof.* Repeat the proof of 8.17, but now only make choices for the odd part lengths  $1, 3, 5, 7$ , etc.  $\square$

**8.20. Theorem: Partitions with Distinct Parts.** Let  $\text{DisPar}$  be the set of integer partitions all of whose parts are distinct. Then

$$\sum_{\mu \in \text{DisPar}} x^{|\mu|} = \prod_{i=1}^{\infty} (1 + x^i).$$

*Proof.* We build a partition  $\mu \in \text{DisPar}$  via the following choice sequence. For each part length  $i \geq 1$ , either choose to not use that part in  $\mu$  or to include that part in  $\mu$  (note that the part is only allowed to occur once). The generating function for this choice is  $1 + x^i$ . The result now follows from the product rule 8.9.  $\square$

By comparing the generating functions in the last two theorems, we are led to the following unexpected result.

**8.21. Theorem: OddPar vs. DisPar.**

$$\sum_{\mu \in \text{OddPar}} x^{|\mu|} = \sum_{\nu \in \text{DisPar}} x^{|\nu|}.$$

*Proof.* We make the following calculation with formal power series:

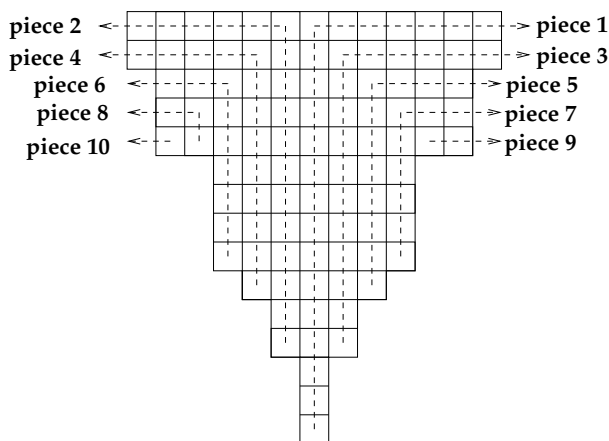
$$\begin{aligned} \sum_{\mu \in \text{OddPar}} x^{|\mu|} &= \prod_{k=1}^{\infty} \frac{1}{1 - x^{2k-1}} \\ &= \prod_{k=1}^{\infty} \frac{1}{1 - x^{2k-1}} \prod_{j=1}^{\infty} \frac{1}{1 - x^{2j}} \prod_{j=1}^{\infty} (1 - x^{2j}) \\ &= \prod_{i=1}^{\infty} \frac{1}{1 - x^i} \prod_{j=1}^{\infty} [(1 - x^j)(1 + x^j)] \\ &= \prod_{i=1}^{\infty} \frac{1}{1 - x^i} \prod_{j=1}^{\infty} (1 - x^j) \prod_{j=1}^{\infty} (1 + x^j) \\ &= \prod_{j=1}^{\infty} (1 + x^j) = \sum_{\nu \in \text{DisPar}} x^{|\nu|}. \end{aligned}$$

The first and last equalities hold by the two preceding theorems. The penultimate equality uses a cancellation of infinite products that was formally justified in 7.42; the same example (with  $x$  replaced by  $x^2$ ) justifies the second equality. The third and fourth equalities can be verified by similar methods; the reader should fill in the details here (cf. 7.140).  $\square$

## 8.6 Partition Bijections

We have just seen that the generating function for partitions into odd parts (relative to area) coincides with the generating function for partitions with distinct parts. We gave an *algebraic proof* of this result based on manipulation of infinite products in the formal power series ring  $\mathbb{Q}[[x]]$ . However, from the combinatorial standpoint, it is natural to ask for a *bijective proof* of the same result. We therefore seek an area-preserving bijection  $F : \text{OddPar} \rightarrow \text{DisPar}$ . Two such bijections are presented in this section.

**8.22. Sylvester's Bijection.** Define  $F : \text{OddPar} \rightarrow \text{DisPar}$  as follows. Given  $\mu \in \text{OddPar}$ , draw a *centered* version of the Ferrers diagram of  $\mu$  in which the middle boxes of the parts of  $\mu$  are all drawn in the same column; see Figure 8.3. Note that each part of  $\mu$  does have a middle box, because the part is odd. Label the columns in the centered diagram of  $\mu$  as  $-k, \dots, -2, -1, 0, 1, 2, \dots, k$  from left to right, so the center column is column 0. Label the rows  $1, 2, 3, \dots$  from top to bottom. We define  $\nu = F(\mu)$  by dissecting the centered diagram



$$F((13,13,11,11,11,7,7,7,7,5,3,3,1,1,1)) = (21,17,16,13,11,9,8,3,2,1)$$

**FIGURE 8.3**

Sylvester's partition bijection.

of  $\mu$  into a sequence of disjoint L-shaped pieces (described below), and letting the parts of  $\nu$  be the number of cells in each piece. The first L-shaped piece consists of all cells in column 0 together with all cells to the right of column 0 in row 1. The second L-shaped piece consists of all cells in column  $-1$  together with all cells left of column  $-1$  in row 1. The third piece consists of the unused cells in column 1 (so row 1 is excluded) together with all cells right of column 1 in row 2. The fourth piece consists of the unused cells in column  $-2$  together with all cells left of column  $-2$  in row 2. We proceed similarly, working outwards in both directions from the center column, cutting off L-shaped pieces that alternately move up and right, then up and left (see Figure 8.3).

One may check geometrically that the size of each L-shaped piece is strictly less than the size of the preceding piece. It follows that  $F(\mu) = \nu = (\nu_1 > \nu_2 > \dots)$  is indeed an element of  $\text{DisPar}$ . Furthermore, since  $|\mu|$  is the sum of the sizes of all the L-shaped pieces, the map  $F : \text{OddPar} \rightarrow \text{DisPar}$  is area-preserving. We must also check that  $F$  is a *bijection* by constructing a map  $G : \text{DisPar} \rightarrow \text{OddPar}$  that is the two-sided inverse of  $F$ .

To see how to define  $G$ , let us examine more closely the dimensions of the L-shaped pieces that appear in the definition of  $F(\mu)$ . Note that each L-shaped piece consists of a corner square, a “vertical portion” of zero or more squares below the corner, and a “horizontal portion” of zero or more squares to the left or right of the corner. Let  $y_0$  be the number of cells in column 0 of the centered diagram of  $\mu$  (so  $y_0 = \ell(\mu)$ ). For all  $i \geq 1$ , let  $x_i$  be the number of cells in the horizontal portion of the  $(2i - 1)$ th L-shaped piece for  $\mu$ . For all  $i \geq 0$ , let  $y_i$  be the number of cells in the vertical portion of the  $2i$ th L-shaped piece for  $\mu$ . For example, in Figure 8.3 we have  $(y_0, y_1, y_2, \dots) = (15, 11, 8, 6, 1, 0, 0, \dots)$  and  $(x_1, x_2, \dots) = (6, 5, 3, 2, 1, 0, 0, \dots)$ . Note that for all  $i \geq 1$ ,  $y_{i-1} > y_i$  whenever  $y_{i-1} > 0$ , and  $x_i > x_{i+1}$  whenever  $x_i > 0$ . Moreover, by the symmetry of the centered diagram of  $\mu$  and the definition of  $F$ , we see that

$$\begin{aligned} \nu_1 &= y_0 + x_1, & \nu_2 &= x_1 + y_1, \\ \nu_3 &= y_1 + x_2, & \nu_4 &= x_2 + y_2, \\ \nu_5 &= y_2 + x_3, & \nu_6 &= x_3 + y_3, \end{aligned}$$

and, in general,

$$\nu_{2i-1} = y_{i-1} + x_i \quad (i \geq 1); \quad \nu_{2i} = x_i + y_i \quad (i \geq 1). \quad (8.2)$$

To compute  $G(\nu)$  for  $\nu \in \text{DisPar}$ , we need to solve the preceding system of equations for  $x_i$  and  $y_i$ , given the part lengths  $\nu_j$ . Noting that  $\nu_k$ ,  $x_k$ , and  $y_k$  must all be zero for large enough indices  $k$ , we can solve for each variable by taking the alternating sum of all the given equations from some point forward. This forces us to define

$$\begin{aligned} y_i &= \nu_{2i+1} - \nu_{2i+2} + \nu_{2i+3} - \nu_{2i+4} + \cdots & (i \geq 0); \\ x_i &= \nu_{2i} - \nu_{2i+1} + \nu_{2i+2} - \nu_{2i+3} + \cdots & (i \geq 1). \end{aligned}$$

One verifies immediately that these choices of  $x_i$  and  $y_i$  do indeed satisfy the equations  $\nu_{2i-1} = y_{i-1} + x_i$  and  $\nu_{2i} = x_i + y_i$ . Furthermore, because the nonzero parts of  $\nu$  are distinct, the required inequalities ( $y_{i-1} > y_i$  whenever  $y_{i-1} > 0$ , and  $x_i > x_{i+1}$  whenever  $x_i > 0$ ) also hold. Now that we know the exact shape of each L-shaped piece, we can fit the pieces together to recover the centered diagram of  $\mu = G(\nu) \in \text{OddPar}$ . For example, given  $\nu = (9, 8, 5, 3, 1, 0, 0, \dots)$ , we compute

$$\begin{aligned} y_0 &= 9 - 8 + 5 - 3 + 1 = 4 \\ x_1 &= 8 - 5 + 3 - 1 = 5 \\ y_1 &= 5 - 3 + 1 = 3 \\ x_2 &= 3 - 1 = 2 \\ y_2 &= 1. \end{aligned}$$

Using this data to reconstitute the centered diagram, we find that  $G(\nu) = (11, 7, 5, 3)$ . In closing, we remark that bijectivity of  $F$  is equivalent to the fact that, for each  $\nu \in \text{DisPar}$ , the system of equations (8.2) has exactly one solution for the unknowns  $x_i$  and  $y_i$ .

**8.23. Glaisher's Bijection.** We define a map  $H : \text{DisPar} \rightarrow \text{OddPar}$  as follows. Each integer  $k \geq 1$  can be written uniquely in the form  $k = 2^e c$ , where  $e \geq 0$  and  $c$  is odd. Given  $\nu \in \text{DisPar}$ , we replace each part  $k$  in  $\nu$  by  $2^e$  copies of the part  $c$  (where  $k = 2^e c$ , as above). Sorting the resulting odd numbers into decreasing order gives us an element  $H(\nu)$  in  $\text{OddPar}$  such that  $|H(\nu)| = |\nu|$ . For example,

$$\begin{aligned} H((15, 12, 10, 8, 6, 3, 1)) &= \text{sort}((15, 3, 3, 3, 3, 5, 5, 1, 1, 1, 1, 1, 1, 1, 3, 3, 3, 1)) \\ &= (15, 5, 5, 3, 3, 3, 3, 3, 3, 3, 1, 1, 1, 1, 1, 1, 1, 1). \end{aligned}$$

The inverse map  $K : \text{OddPar} \rightarrow \text{DisPar}$  is defined as follows. Consider a partition  $\mu \in \text{OddPar}$ . For each odd number  $c$  that appears as a part of  $\mu$ , let  $n = n(c) \geq 1$  be the number of times  $c$  occurs in  $\mu$ . We can write  $n$  uniquely as a sum of distinct powers of 2 (this is the base-2 expansion of the integer  $n$ , cf. 5.5). Say  $n = 2^{d_1} + 2^{d_2} + \cdots + 2^{d_s}$ . We replace the  $n$  copies of  $c$  in  $\mu$  by parts of size  $2^{d_1}c$ ,  $2^{d_2}c$ ,  $\dots$ ,  $2^{d_s}c$ . These parts are distinct from one another (since the  $d_j$ 's are distinct), and they are also distinct from the parts obtained in the same way from other odd values of  $c$  appearing as parts of  $\mu$ . Sorting the parts thus gives a partition  $K(\mu) \in \text{DisPar}$ . For example,

$$K((7, 7, 7, 7, 3, 3, 3, 3, 3, 1, 1, 1)) = \text{sort}((28, 7, 12, 6, 2, 1)) = (28, 12, 7, 6, 2, 1).$$

It is readily verified that  $H \circ K$  and  $K \circ H$  are identity maps.

Glaisher's bijection generalizes to prove the following theorem.

**8.24. Theorem: Glaisher's Partition Identity.** For all  $d \geq 2$  and  $N \geq 0$ , the number of partitions of  $N$  where no part repeats  $d$  or more times equals the number of partitions of  $N$  with no part divisible by  $d$ .

*Proof.* For fixed  $d$ , let  $A$  be the set of partitions where no part repeats  $d$  or more times, and let  $B$  be the set of partitions with no part divisible by  $d$ . It suffices to describe weight-preserving maps  $H : A \rightarrow B$  and  $K : B \rightarrow A$  such that  $H \circ K$  and  $K \circ H$  are identity maps. We define  $K$  by analogy with what we did above. Fix  $\mu \in B$ . For each  $c$  that appears as a part of  $\mu$ , let  $n = n(c)$  be the number of times this part occurs in  $\mu$ . Write  $n$  in base  $d$  as

$$n = \sum_{k=0}^s a_k d^k \quad (0 \leq a_k < d),$$

where  $n$  and  $a_0, \dots, a_s$  all depend on  $c$ . To construct  $K(\mu)$ , we replace the  $n$  copies of  $c$  in  $\mu$  by  $a_0$  copies of  $d^0 c$ ,  $a_1$  copies of  $d^1 c$ ,  $\dots$ ,  $a_k$  copies of  $d^k c$ ,  $\dots$ , and  $a_s$  copies of  $d^s c$ . One checks that the resulting partition lies in  $A$ , using the fact that no part  $c$  of  $\mu$  is divisible by  $d$ .

To compute  $H(\nu)$  for  $\nu \in A$ , note that each part  $m$  in  $\nu$  can be written uniquely in the form  $m = d^k c$  for some  $k \geq 0$  and some  $c = c(m)$  not divisible by  $d$ . Adding up all such parts of  $\nu$  that have the same value of  $c$  produces an expression of the form  $\sum_{k \geq 0} a_k d^k c$ , where  $0 \leq a_k < d$  by definition of  $A$ . To get  $H(\nu)$ , we replace all these parts by  $\sum_{k \geq 0} a_k d^k$  copies of the part  $c$ , for every possible  $c$  not divisible by  $d$ . Comparing the descriptions of  $H$  and  $K$ , one sees that these two maps are inverses.  $\square$

**8.25. Remark: Rogers-Ramanujan Identities.** A huge number of partition identities have been discovered, which are similar in character to the one we just proved. Two especially famous examples are the *Rogers-Ramanujan identities*. The first such identity says that, for all  $N$ , the number of partitions of  $N$  into parts congruent to 1 or 4 modulo 5 equals the number of partitions of  $N$  into distinct parts  $\nu_1 > \nu_2 > \dots > \nu_k > 0$  such that  $\nu_i - \nu_{i+1} \geq 2$  for all  $i < k$ . The second identity says that, for all  $N$ , the number of partitions of  $N$  into parts congruent to 2 or 3 modulo 5 equals the number of partitions of  $N$  into distinct parts  $\nu_1 > \nu_2 > \dots > \nu_k > 0 = \nu_{k+1}$  such that  $\nu_i - \nu_{i+1} \geq 2$  for all  $i \leq k$ . One can seek algebraic and/or bijective proofs for these and other identities. Proofs of both types are known for the Rogers-Ramanujan identities, but the bijective proofs are all quite complicated.

## 8.7 Euler's Pentagonal Number Theorem

We have seen that  $\prod_{i \geq 1} (1 + x^i)$  is the generating function for partitions with distinct parts, whereas  $\prod_{i \geq 1} (1 - x^i)^{-1}$  is the generating function for all integer partitions. This section investigates the infinite product  $\prod_{i \geq 1} (1 - x^i)$ , which is the multiplicative inverse for the partition generating function (see 7.42). The next theorem shows that expanding this product leads to a remarkable amount of cancellation of terms due to the minus signs (cf. 7.139).

**8.26. Pentagonal Number Theorem.**

$$\begin{aligned} \prod_{i=1}^{\infty} (1 - x^i) &= 1 + \sum_{n=1}^{\infty} (-1)^n [x^{n(3n-1)/2} + x^{n(3n+1)/2}] \\ &= 1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + x^{22} + x^{26} - x^{35} - x^{40} + \dots \end{aligned}$$

*Proof.* Consider the set DisPar of integer partitions with distinct parts, weighted by area. For  $\mu \in \text{DisPar}$ , define the *sign* of  $\mu$  to be  $(-1)^{\ell(\mu)}$ . By modifying the argument in 8.20 to include these signs, we obtain

$$\prod_{i=1}^{\infty} (1 - x^i) = \sum_{\mu \in \text{DisPar}} (-1)^{\ell(\mu)} x^{|\mu|}.$$

We now define an ingenious area-preserving, sign-reversing involution  $I$  on DisPar (due to Franklin). Given a partition  $\mu = (\mu_1 > \mu_2 > \cdots > \mu_s) \in \text{DisPar}$ , let  $a \geq 1$  be the largest index such that the part sizes  $\mu_1, \mu_2, \dots, \mu_a$  are consecutive integers, and let  $b = \mu_s$  be the smallest part of  $\mu$ . Figure 8.4 shows how  $a$  and  $b$  can be read from the Ferrers diagram of  $\mu$ . For most partitions  $\mu$ , we define  $I$  as follows. If  $a < b$ , let  $I(\mu)$  be the partition obtained by decreasing the first  $a$  parts of  $\mu$  by 1 and adding a new part of size  $a$  to the end of  $\mu$ . If  $a \geq b$ , let  $I(\mu)$  be the partition obtained by removing the last part of  $\mu$  (of size  $b$ ) and increasing the first  $b$  parts of  $\mu$  by 1 each. See the examples in Figure 8.4.  $I$  is weight-preserving and sign-reversing, since  $I(\mu)$  has either one more or one fewer part than  $\mu$ . It is also routine to check that  $I(I(\mu)) = \mu$ . Thus we can cancel out all the pairs of objects  $\{\mu, I(\mu)\}$ .

It may seem at first glance that we have canceled *all* the objects in DisPar! However, there are some choices of  $\mu$  where the definition of  $I(\mu)$  in the previous paragraph fails to produce a partition with distinct parts. Consider what happens in the “overlapping” situation  $a = \ell(\mu)$ . If  $b = a + 1$  in this situation, the prescription for creating  $I(\mu)$  leads to a partition whose smallest two parts both equal  $a$ . On the other hand, if  $b = a$ , the definition of  $I(\mu)$  fails because there are not enough parts left to increment by 1 after dropping the smallest part of  $\mu$ . In all other cases, the definition of  $I$  works even when  $a = \ell(\mu)$ . We see now that there are two classes of partitions that cannot be canceled by  $I$  (Figure 8.5). First, there are partitions of the form  $(2n, 2n - 1, \dots, n + 1)$ , which have length  $n$  and area  $n(3n + 1)/2$ , for all  $n \geq 1$ . Second, there are partitions of the form  $(2n - 1, 2n - 2, \dots, n)$ , which have length  $n$  and area  $n(3n - 1)/2$ , for all  $n \geq 1$ . Furthermore, the empty partition is not canceled by  $I$ . Adding up these signed, weighted objects gives the right side of the equation in the theorem.  $\square$

We can now deduce Euler’s recursion for counting integer partitions that we stated in 2.48.

**8.27. Theorem: Partition Recursion.** For every  $n \in \mathbb{Z}$ , let  $p(n)$  be the number of integer partitions of  $n$ . The numbers  $p(n)$  satisfy the recursion

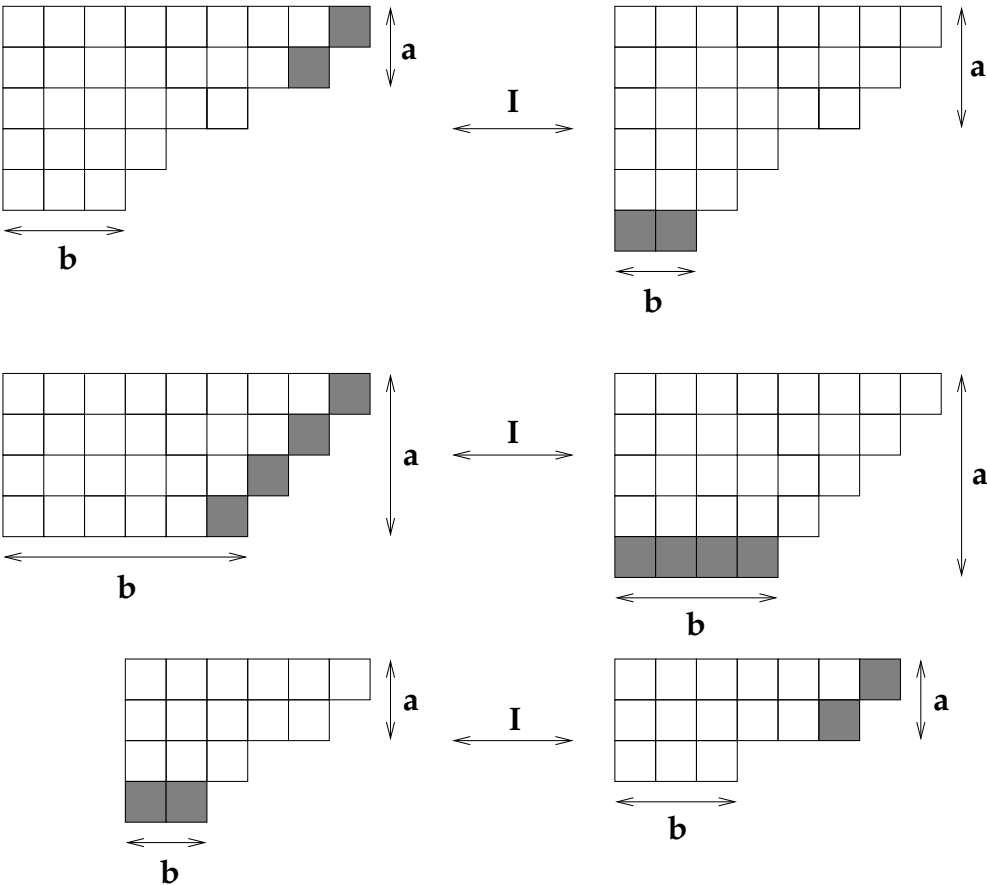
$$\begin{aligned} p(n) &= p(n - 1) + p(n - 2) - p(n - 5) - p(n - 7) + p(n - 12) + p(n - 15) - \cdots \\ &= \sum_{k \geq 1} (-1)^{k-1} [p(n - k(3k - 1)/2) + p(n - k(3k + 1)/2)] \end{aligned} \tag{8.3}$$

for  $n \geq 1$ . The initial conditions are  $p(0) = 1$  and  $p(n) = 0$  for all  $n < 0$ .

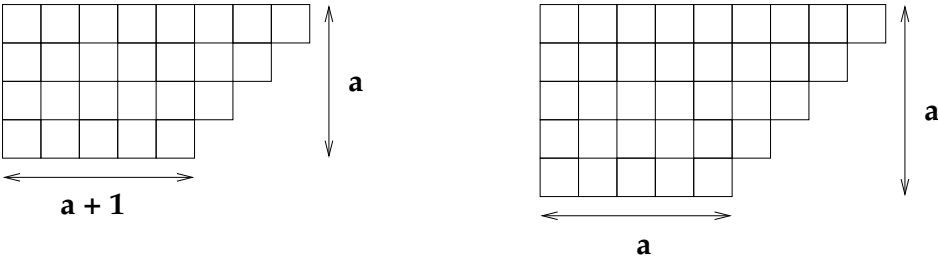
*Proof.* We have proved the identities

$$\begin{aligned} \prod_{i \geq 1} \frac{1}{1 - x^i} &= \sum_{\mu \in \text{Par}} x^{|\mu|} = \sum_{n \geq 0} p(n) x^n, \\ \prod_{i \geq 1} (1 - x^i) &= 1 + \sum_{k \geq 1} (-1)^k [x^{k(3k-1)/2} + x^{k(3k+1)/2}]. \end{aligned}$$

The product of the left sides of these two identities is 1, so the product of the right sides is



**FIGURE 8.4**  
Franklin's partition involution.



**FIGURE 8.5**  
Fixed points of Franklin's involution.



also 1. Thus, for each  $n \geq 1$ , the coefficient of  $x^n$  in the product

$$\left( \sum_{n \geq 0} p(n) x^n \right) \cdot \left( 1 + \sum_{k \geq 1} (-1)^k [x^{k(3k-1)/2} + x^{k(3k+1)/2}] \right)$$

is zero. This coefficient also equals  $p(n) - p(n-1) - p(n-2) + p(n-5) + p(n-7) - \dots$ . Solving for  $p(n)$  yields the recursion in the theorem.  $\square$

## 8.8 Stirling Numbers of the First Kind

We can often translate combinatorial recursions into generating functions for the objects in question. We illustrate this process by developing generating functions for the Stirling numbers of the first and second kind.

Recall from §3.6 that the signless Stirling number of the first kind (which we denote here by  $c(n, k)$ ) counts the number of permutations of  $n$  objects whose functional digraphs consist of  $k$  disjoint cycles. These numbers satisfy  $c(n, 0) = 0$  for  $n > 0$ ,  $c(n, n) = 1$  for  $n \geq 0$ , and

$$c(n, k) = c(n-1, k-1) + (n-1)c(n-1, k) \quad (0 < k < n).$$

We also set  $c(n, k) = 0$  whenever  $k < 0$  or  $k > n$ .

Define a polynomial  $f_n = \sum_{k=0}^n c(n, k) t^k \in \mathbb{Q}[t]$  for each  $n \geq 0$ , and define the formal power series

$$F = \sum_{n \geq 0} \frac{f_n}{n!} x^n = \sum_{n \geq 0} \sum_{k=0}^n \frac{c(n, k)}{n!} t^k x^n \in \mathbb{Q}(t)[[x]].$$

As we will see, it is technically convenient to introduce the denominators  $n!$  as done here. Note that the coefficient of  $t^k x^n$  in  $F$ , namely  $c(n, k)/n!$ , is the *probability* that a randomly chosen permutation of  $n$  objects will have  $k$  cycles. The coefficient of  $x^0$  in  $F$  is the polynomial  $1 \in \mathbb{Q}[t]$ .

We will use the recursion for  $c(n, k)$  to prove the relation  $(1-x)D_x F = tF$ , where  $D_x F$  is the formal partial derivative of  $F$  with respect to  $x$  (§7.16). We first compute

$$\begin{aligned} D_x F &= \sum_{n \geq 0} \sum_{k=0}^n \frac{nc(n, k)}{n!} t^k x^{n-1} = \sum_{n \geq 0} \sum_{k=0}^n \frac{n[c(n-1, k-1) + (n-1)c(n-1, k)]}{n!} t^k x^{n-1} \\ &= \sum_{n \geq 0} \sum_{k=0}^n \frac{c(n-1, k-1)}{(n-1)!} t^k x^{n-1} + \sum_{n \geq 0} \sum_{k=0}^n \frac{c(n-1, k)}{(n-2)!} t^k x^{n-1}. \end{aligned}$$

In the first summation, let  $m = n-1$  and  $j = k-1$ . After discarding zero terms, we see that

$$\sum_{n \geq 0} \sum_{k=0}^n \frac{c(n-1, k-1)}{(n-1)!} t^k x^{n-1} = t \sum_{m \geq 0} \sum_{j=0}^m \frac{c(m, j)}{m!} t^j x^m = tF.$$

On the other hand, letting  $m = n-1$  in the second summation shows that

$$\sum_{n \geq 0} \sum_{k=0}^n \frac{c(n-1, k)}{(n-2)!} t^k x^{n-1} = x \sum_{m \geq 0} \sum_{k=0}^m c(m, k) t^k \frac{x^{m-1}}{(m-1)!} = x D_x F,$$

since  $D_x(x^m/m!) = x^{m-1}/(m-1)!$ . So we indeed have  $D_x F = tF + xD_x F$ , as claimed.

We now know a formal differential equation satisfied by  $F$ , together with the initial condition  $F(0) = 1$ . To find an explicit formula for  $F$ , we need only “solve for  $F$ ” by techniques that may be familiar from calculus (cf. 7.176). However, one must remember that all our computations need to be justifiable at the level of *formal* power series. Let us work in the ring  $R = K[[x]]$  where  $K$  is the field  $\mathbb{Q}(t)$ . We begin by writing the differential equation in the form

$$(1 - x)D_x F = tF.$$

Since  $1 - x \in R$  and  $F \in R$  have nonzero constant terms, we can divide by these quantities (using 7.40) to arrive at

$$\frac{D_x F}{F} = \frac{t}{1 - x}.$$

Now,  $\log(F)$  is defined because  $F(0) = 1$  (see 7.92), and the derivative of  $\log(F)$  is  $(D_x F)/F$  by 7.96. On the other hand, using 7.96 and 7.79, we see that  $\frac{t}{1-x}$  is the formal derivative of  $\log[(1 - x)^{-t}]$ . We therefore have

$$\frac{d}{dx} (\log(F)) = \frac{d}{dx} (\log[(1 - x)^{-t}]).$$

Since both logarithms have constant term zero, we deduce

$$\log(F) = \log[(1 - x)^{-t}]$$

using 7.130 or 7.132. Finally, taking the formal exponential of both sides and using 7.94 gives

$$F = (1 - x)^{-t}. \quad (8.4)$$

Having discovered this formula for  $F$ , we can now give an independent verification of its correctness by invoking our earlier results on Stirling numbers and generalized powers:

$$\begin{aligned} (1 - x)^{-t} &= \sum_{n \geq 0} \frac{(-t) \downarrow_n}{n!} (-x)^n \quad \text{by 7.74} \\ &= \sum_{n \geq 0} \frac{(t) \uparrow_n}{n!} x^n \quad \text{by 2.76} \\ &= \sum_{n \geq 0} f_n \frac{x^n}{n!} \quad \text{by 2.78} \\ &= F. \end{aligned}$$

## 8.9 Stirling Numbers of the Second Kind

In this section, we derive a generating function for Stirling numbers of the second kind. Recall from §2.9 that  $S(n, k)$  is the number of set partitions of an  $n$ -element set into  $k$  nonempty blocks. We will study the formal power series

$$G = \sum_{n \geq 0} \sum_{k=0}^n \frac{S(n, k)}{n!} t^k x^n \in \mathbb{Q}(t)[[x]].$$

The following recursion will help us find a differential equation satisfied by  $G$ .

**8.28. Theorem: Recursion for Stirling Numbers.** For all  $n \geq 0$  and  $0 \leq k \leq n+1$ ,

$$S(n+1, k) = \sum_{i=0}^n \binom{n}{i} S(n-i, k-1).$$

The initial conditions are  $S(0, 0) = 1$  and  $S(n, k) = 0$  whenever  $k < 0$  or  $k > n$ .

*Proof.* Consider set partitions of  $\{1, 2, \dots, n+1\}$  into  $k$  blocks such that the block containing  $n+1$  has  $i$  other elements in it (where  $0 \leq i \leq n$ ). To build such a set partition, choose the  $i$  elements that go in the block with  $n+1$  in  $\binom{n}{i}$  ways, and then choose a set partition of the remaining  $n-i$  elements into  $k-1$  blocks. The recursion now follows from the sum and product rules. (Compare to the proof of 2.53.)  $\square$

**8.29. Theorem: Differential Equation for  $G$ .** The series  $G = \sum_{n \geq 0} \sum_{k=0}^n \frac{S(n, k)}{n!} t^k x^n$  satisfies  $G(0) = 1$  and

$$D_x G = t e^x G.$$

*Proof.* The derivative of  $G$  with respect to  $x$  is

$$D_x G = \sum_{m \geq 0} \sum_{k=0}^m \frac{S(m, k)}{m!} t^k m x^{m-1} = \sum_{n \geq 0} \sum_{k=0}^{n+1} \frac{S(n+1, k)}{n!} t^k x^n,$$

where we have set  $n = m - 1$ . Using 8.28 transforms this expression into

$$\sum_{n \geq 0} \sum_{k=0}^{n+1} \sum_{i=0}^n \frac{1}{n!} \binom{n}{i} S(n-i, k-1) t^k x^n = \sum_{n \geq 0} \sum_{k=0}^{n+1} \sum_{i=0}^n \frac{S(n-i, k-1) t^k x^{n-i}}{(n-i)!} \cdot \frac{x^i}{i!}.$$

Setting  $j = k - 1$ , the formula becomes

$$t \sum_{n \geq 0} \sum_{j=0}^n \sum_{i=0}^n \frac{S(n-i, j) t^j x^{n-i}}{(n-i)!} \cdot \frac{x^i}{i!} = t \sum_{n \geq 0} \sum_{i=0}^n \sum_{j=0}^{n-i} \frac{S(n-i, j) t^j x^{n-i}}{(n-i)!} \cdot \frac{x^i}{i!}.$$

Finally, recalling 7.6, the last expression equals

$$t \left( \sum_{i \geq 0} \frac{x^i}{i!} \right) \cdot \left( \sum_{m \geq 0} \left[ \sum_{j=0}^m \frac{S(m, j)}{m!} t^j \right] x^m \right) = t e^x G. \quad \square$$

**8.30. Theorem: Generating Function for Stirling Numbers of the Second Kind.**

$$\sum_{n \geq 0} \sum_{k=0}^n \frac{S(n, k)}{n!} t^k x^n = \exp[t(e^x - 1)] \in \mathbb{Q}(t)[[x]].$$

*Proof.* Call the left side  $G$ , as above. We proceed to solve the differential equation  $D_x G = t e^x G$  and initial condition  $G(0) = 1$ . The series  $G$  is invertible, having nonzero constant term, so we have  $(D_x G)/G = t e^x$ . Formally integrating both sides with respect to  $x$  (cf. 7.130) leads to  $\log(G) = t e^x + c$ , where  $c \in \mathbb{Q}(t)$ . As  $\log(G)_0 = 0$  and  $(t e^x)_0 = t$ , the constant of integration must be  $c = -t$ . Finally, exponentiating both sides shows that

$$G = \exp(t e^x - t) = \exp[t(e^x - 1)]. \quad \square$$

**8.31. Theorem: Generating Function for  $S(n, k)$  for fixed  $k$ .** For all  $k \geq 0$ ,

$$\sum_{n \geq k} S(n, k) \frac{x^n}{n!} = \frac{1}{k!} (e^x - 1)^k \in \mathbb{Q}[[x]].$$

*Proof.* We have

$$\exp[t(e^x - 1)] = \sum_{k \geq 0} t^k \frac{(e^x - 1)^k}{k!} \in \mathbb{Q}((x))[[t]].$$

Extracting the coefficient of  $t^k$  and using 8.30, we obtain the desired formula.  $\square$

## 8.10 The Exponential Formula

Many combinatorial structures can be decomposed into disjoint unions of “connected components.” For example, set partitions consist of a collection of disjoint blocks; permutations can be regarded as a collection of disjoint cycles; and graphs are disjoint unions of connected graphs. The *exponential formula* allows us to compute the generating function for such structures from the generating function for the connected “building blocks” of which they are composed.

For each  $k \geq 1$ , let  $\mathcal{C}_k$  be an admissible weighted set of “connected structures of size  $k$ ,” and let  $C_k = \sum_{z \in \mathcal{C}_k} \text{wt}(z) \in \mathbb{Q}[[t]]$ . We introduce the generating function  $C^* = \sum_{k \geq 1} \frac{C_k}{k!} x^k$  to encode information about all the sets  $\mathcal{C}_k$  (cf. 7.168). Next, we must formally define a set of structures of size  $n$  that consist of disjoint unions of connected structures of various sizes summing to  $n$ . For every  $n \geq 0$ , let  $U_n$  be the set of pairs  $(S, f)$  such that:  $S$  is a set partition of  $\{1, 2, \dots, n\}$ , and  $f : S \rightarrow \bigcup_{k \geq 1} \mathcal{C}_k$  is a function such that  $f(A) \in \mathcal{C}_{|A|}$  for all  $A \in S$ . This says that for every  $m$ -element block  $A$  of the set partition  $S$ ,  $f(A)$  is a connected structure of size  $m$ . Let  $\text{wt}(S, f) = \prod_{A \in S} \text{wt}(f(A))$ , and define

$$F = \sum_{n \geq 0} \left( \sum_{u \in U_n} \text{wt}(u) \right) \frac{x^n}{n!}.$$

Note that  $F(0) = 1$ .

**8.32. Theorem: Exponential Formula.** With notation as above,  $F = \exp(C^*)$ .

*Proof.* By 7.92,

$$\exp(C^*) = \sum_{m \geq 0} \frac{1}{m!} \left( \sum_{k \geq 1} \frac{C_k}{k!} x^k \right)^m.$$

This series has constant term 1. For  $n > 0$ , the coefficient of  $x^n$  in  $\exp(C^*)$  is (by 7.10)

$$\sum_{m=1}^n \frac{1}{m!} \sum_{k_1=1}^n \sum_{k_2=1}^n \cdots \sum_{k_m=1}^n \chi(k_1 + k_2 + \cdots + k_m = n) \frac{C_{k_1} C_{k_2} \cdots C_{k_m}}{k_1! k_2! \cdots k_m!}.$$

This coefficient can also be written

$$\frac{1}{n!} \sum_{m=1}^n \frac{1}{m!} \sum_{\substack{(k_1, \dots, k_m): \\ k_i > 0, k_1 + \cdots + k_m = n}} \binom{n}{k_1, k_2, \dots, k_m} C_{k_1} C_{k_2} \cdots C_{k_m}.$$

Comparing to the definition of  $F$ , we need to prove that

$$\sum_{u \in U_n} \text{wt}(u) = \sum_{m=1}^n \frac{1}{m!} \sum_{\substack{(k_1, \dots, k_m): \\ k_i > 0, k_1 + \dots + k_m = n}} \binom{n}{k_1, k_2, \dots, k_m} C_{k_1} C_{k_2} \cdots C_{k_m}.$$

Consider objects  $(S, f) \in U_n$  for which  $|S| = m$  (so that the object has  $m$  “connected components”). By the sum rule, it will suffice to prove

$$m! \sum_{\substack{(S, f) \in U_n: \\ |S| = m}} \text{wt}(u) = \sum_{\substack{(k_1, \dots, k_m): \\ k_i > 0, k_1 + \dots + k_m = n}} \binom{n}{k_1, k_2, \dots, k_m} C_{k_1} C_{k_2} \cdots C_{k_m} \quad (1 \leq m \leq n).$$

The left side counts pairs  $(T, f)$ , where  $T = (T_1, T_2, \dots, T_m)$  is an *ordered* set partition of  $\{1, 2, \dots, n\}$  into  $m$  blocks, and for each  $i \leq m$ ,  $f(T_i) \in \mathcal{C}_{|T_i|}$ . (We must multiply by  $m!$  to pass from the set partition  $\{T_1, \dots, T_m\}$  to the ordered set partition  $(T_1, \dots, T_m)$ .) The right side counts the same set of objects, as we see by the following counting argument. Let  $(k_1, \dots, k_m)$  be the sizes of the blocks in the ordered list  $T = (T_1, \dots, T_m)$ . We can identify  $T$  with a word  $w = w_1 w_2 \cdots w_n$  in  $\mathcal{R}(1^{k_1} \cdots m^{k_m})$  by letting  $w_i = j$  iff  $i \in T_j$ . It follows that there are  $\binom{n}{k_1, k_2, \dots, k_m}$  choices for the ordered set partition  $T$  (see 1.46). Next, for  $1 \leq i \leq m$ , we choose  $f(T_i) \in \mathcal{C}_{k_i}$ . The generating function for this choice is  $C_{k_i}$ . The formula on the right side now follows from the product and sum rules.  $\square$

The generating functions for Stirling numbers (derived in §8.8 and §8.9) are consequences of the exponential formula, as we now show.

**8.33. Example: Bell Numbers and Stirling Numbers of the Second Kind.** For each  $k \geq 1$ , let  $\mathcal{C}_k$  consist of a single element of weight 1. Then  $C^* = \sum_{k \geq 1} 1x^k/k! = e^x - 1$ . With this choice of the sets  $\mathcal{C}_k$ , an element  $(S, f) \in U_n$  can be identified with the set partition  $S$  of an  $n$ -element set, since there is only one possible choice for the function  $f$ . Therefore, in this example,

$$F = \sum_{u \in U_n} \text{wt}(u) \frac{x^n}{n!} = \sum_{n \geq 0} \frac{B(n)}{n!} x^n,$$

where the Bell number  $B(n)$  counts set partitions (see 2.51). The exponential formula now gives

$$\sum_{n \geq 0} \frac{B(n)}{n!} x^n = \exp(e^x - 1).$$

Intuitively, the unique element in  $\mathcal{C}_k$  is the  $k$ -element set  $\{1, 2, \dots, k\}$  which is the prototypical example of a  $k$ -element block in a set partition. If we let this element have weight  $t$  (for all  $k$ ), then  $\text{wt}(S, f) = t^{|S|}$  will encode the number of blocks in the set partition  $S$ . In this case,  $C^* = t(e^x - 1)$  and the exponential formula gives

$$\sum_{n \geq 0} \sum_{k=0}^n \frac{S(n, k)}{n!} t^k x^n = \exp[t(e^x - 1)],$$

in agreement with 8.30.

**8.34. Example: Stirling Numbers of the First Kind.** For each  $k \geq 1$ , let  $\mathcal{C}_k$  consist

of all  $k$ -cycles on  $\{1, 2, \dots, k\}$ , each having weight  $t$ . Since there are  $(k-1)!$  such  $k$ -cycles, we have

$$C^* = \sum_{k \geq 1} \frac{t(k-1)!}{k!} x^k = t \sum_{k \geq 1} \frac{x^k}{k} = -t \log(1-x) = \log[(1-x)^{-t}]$$

(the last step used 7.97). Consider an element  $(S, f) \in U_n$ . If  $A = \{i_1 < i_2 < \dots < i_k\}$  is a block of  $S$ , then  $f(A)$  is some  $k$ -cycle  $(j_1, j_2, \dots, j_k)$ , where  $j_1, \dots, j_k$  is a rearrangement of  $1, 2, \dots, k$ . By replacing the numbers  $1, 2, \dots, k$  in this  $k$ -cycle by  $i_1, i_2, \dots, i_k$ , we obtain a  $k$ -cycle with vertex set  $A$ . Doing this for every  $A \in S$  produces the functional digraph of a permutation of  $n$  elements. More formally, we have just defined a bijection from  $U_n$  to the set  $S_n$  of permutations of  $\{1, 2, \dots, n\}$ . Note that  $\text{wt}(S, f) = t^{|S|} = t^c$ , where  $c$  is the number of cycles in the permutation associated to  $(S, f)$ . It follows from these observations and the exponential formula that

$$\sum_{n \geq 0} \sum_{k=0}^n \frac{c(n, k)}{n!} t^k x^n = \sum_{n \geq 0} \sum_{w \in S_n} \frac{1}{n!} t^{\text{cyc}(w)} x^n = \exp(C^*) = (1-x)^{-t},$$

in agreement with (8.4).

In the next example, we use the inverse of the exponential formula to deduce the generating function  $C^*$  from knowledge of the generating function  $F$ .

**8.35. Example: Connected Components of Graphs.** For each  $k \geq 1$ , let  $\mathcal{C}_k$  consist of all *connected* simple graphs on the vertex set  $\{1, 2, \dots, k\}$ ; let each such graph have weight 1. Direct computation of the generating function  $C^*$  is difficult. On the other hand, consider an object  $(S, f) \in U_n$ . Given a block  $A = \{i_1 < i_2 < \dots < i_k\}$  in  $S$ ,  $f(A)$  is a connected graph with vertex set  $1, 2, \dots, k$ . Renaming these vertices to be  $i_1, i_2, \dots, i_k$  produces a connected graph with vertex set  $A$ . Doing this for every  $A \in S$  produces an *arbitrary* simple graph with vertex set  $\{1, 2, \dots, n\}$ . Thus,  $F$  is the generating function for such graphs, with  $x$  keeping track of the number of vertices. There are  $2^{\binom{n}{2}}$  simple graphs on  $n$  vertices, since we can either include or exclude each of the  $\binom{n}{2}$  possible edges. Accordingly,

$$\exp(C^*) = F = \sum_{n \geq 0} \frac{2^{\binom{n}{2}}}{n!} x^n$$

and so  $C^* = \log(F)$ . Extracting the coefficient of  $x^n/n!$  on both sides leads to the exact formula

$$\sum_{m=1}^n \sum_{\substack{(k_1, \dots, k_m): \\ k_i > 0, k_1 + \dots + k_m = n}} \frac{(-1)^{m-1}}{m} \binom{n}{k_1, \dots, k_m} 2^{\binom{k_1}{2} + \dots + \binom{k_m}{2}} \quad (8.5)$$

for the number of connected simple graphs on  $n$  vertices.

---

## Summary

- *Admissible Sets and Generating Functions.* A weighted set  $(T, \text{wt})$  is admissible iff  $T_n = \{z \in T : \text{wt}(z) = n\}$  is finite for all  $n \geq 0$ . The generating function for an admissible weighted set is  $G_{T, \text{wt}} = \sum_{n=0}^{\infty} |T_n| x^n \in \mathbb{Q}[[x]]$ . Two weighted sets have the same generating function iff there is a weight-preserving bijection between them.

- *Sum Rule for Generating Functions.* If an admissible weighted set  $S$  is a finite or countable disjoint union of subsets  $T_i$ , then each  $T_i$  is admissible and  $G_S = \sum_i G_{T_i}$ .
- *Finite Product Rule for Generating Functions.* If  $T = T_1 \times \cdots \times T_k$  is a finite product of admissible weighted sets, with  $\text{wt}(z_1, \dots, z_k) = \sum_i \text{wt}(z_i)$ , then  $T$  is admissible and  $G_T = \prod_{i=1}^k G_{T_i}$ .
- *Infinite Product Rule for Generating Functions.* Suppose  $\{T_n : n \geq 1\}$  is a family of admissible weighted sets where each  $T_n$  has a unique element  $1_n$  of weight zero and  $\text{ord}(G_{T_n} - 1) \rightarrow \infty$  as  $n \rightarrow \infty$ . Let  $T = \prod_{n \geq 1}^* T_n$  be the set of sequences  $z = (z_n : n \geq 1)$  with  $z_n \in T_n$  and  $z_n = 1_n$  for all large enough  $n$ , with  $\text{wt}(z) = \sum_n \text{wt}(z_n)$ . Then  $T$  is admissible and  $G_T = \prod_{n=1}^{\infty} G_{T_n}$ .
- *Generating Functions for Trees.* The formal series

$$\frac{1 - \sqrt{1 - 4x}}{2} = \sum_{n \geq 1} C_{n-1} x^n = \sum_{n \geq 1} \frac{1}{2n-1} \binom{2n-1}{n-1, n}$$

is the generating function for the following sets of weighted trees: (a) binary trees, weighted by number of vertices plus 1; (b) nonempty full binary trees, weighted by number of leaves; (c) ordered trees, weighted by number of vertices.

- *Compositional Inversion Formulas.* Given  $F = x/R$  where  $R \in K[[x]]$  and  $R(0) \neq 0$ , the unique formal series  $G$  such that  $F \bullet G = x = G \bullet F$  is the generating function for the set of terms, where the weight of a term  $w_1 \cdots w_n$  is  $x^n R_{w_1} \cdots R_{w_n}$ . Furthermore,  $G(n) = (R^n)_{n-1}/n = [(d/dx)^{n-1} R^n]_0/n!$  for  $n \geq 1$ .
- *Partition Generating Functions.* By building integer partitions row by row or column by column and using the product rule, one sees that

$$\begin{aligned} \sum_{\mu \in \text{Par}} t^{\ell(\mu)} x^{|\mu|} &= \prod_{i=1}^{\infty} \frac{1}{1 - tx^i} = \sum_{\mu \in \text{Par}} t^{\mu_1} x^{|\mu|}; \\ \sum_{\mu \in \text{OddPar}} x^{|\mu|} &= \prod_{k=1}^{\infty} \frac{1}{1 - x^{2k-1}} = \prod_{i=1}^{\infty} (1 + x^i) = \sum_{\mu \in \text{DisPar}} x^{|\mu|}. \end{aligned}$$

Sylvester's bijection dissects the centered Ferrers diagram of  $\mu \in \text{OddPar}$  into  $L$ -shaped pieces that give a partition in  $\text{DisPar}$ . Glaisher's bijection replaces each part  $k = 2^e c$  in a partition  $\nu \in \text{DisPar}$  (where  $e \geq 0$  and  $c$  is odd) by  $2^e$  copies of  $c$ , giving a partition in  $\text{OddPar}$ .

- *Pentagonal Number Theorem.* Franklin proved

$$\prod_{i=1}^{\infty} (1 - x^i) = 1 + \sum_{n=1}^{\infty} (-1)^n [x^{n(3n-1)/2} + x^{n(3n+1)/2}]$$

by an involution on signed partitions with distinct parts. The map moves boxes between the “staircase” at the top of the partition and the bottom row; this move cancels all partitions except the “pentagonal” ones counted by the right side. Since  $\prod_{i \geq 1} (1 - x^i)$  is the inverse of the partition generating function, we deduce the partition recursion

$$p(n) = \sum_{k \geq 1} (-1)^{k-1} [p(n - k(3k-1)/2) + p(n - k(3k+1)/2)].$$

- *Generating Functions for Stirling Numbers.* By solving formal differential equations or using the exponential formula, one obtains

$$\sum_{n \geq 0} \sum_{k=0}^n \frac{c(n, k)}{n!} t^k x^n = (1 - x)^{-t}; \quad \sum_{n \geq 0} \sum_{k=0}^n \frac{S(n, k)}{n!} t^k x^n = \exp[t(e^x - 1)].$$

Hence,  $\sum_{n \geq k} S(n, k) x^n / n! = (e^x - 1)^k / k!$  for  $k \geq 0$ .

- *Exponential Formula.* Suppose  $T_k$  is a weighted set with  $G_{T_k} \in \mathbb{Q}[[t]]$ , for  $k \geq 0$ . Let  $C^* = \sum_{k \geq 1} (x^k / k!) G_{T_k} \in \mathbb{Q}[[t, x]]$ , and let  $U_n$  be the set of pairs  $(S, f)$  where  $S$  is a set partition of  $\{1, 2, \dots, n\}$  and  $f$  is a function on  $S$  with  $f(A) \in T_{|A|}$  for all  $A \in S$ . Let  $\text{wt}(S, f) = \prod_{A \in S} \text{wt}(f(A))$  and  $F = \sum_{n \geq 0} \sum_{(S, f) \in U_n} \text{wt}(S, f) x^n / n!$ . Then  $F = \exp(C^*)$ . Informally, if  $C^*$  is the exponential generating function for a set of connected building blocks, then  $F = \exp(C^*)$  is the exponential generating function for the set of objects obtained by taking labeled disjoint unions of these building blocks.

## Exercises

**8.36.** (a) Let  $W$  be the set of all words over a  $k$ -letter alphabet, weighted by length. Find the generating function  $G_W$ . (b) Show that  $x(G'_W)$  is the generating function for the set of all nonempty words in which one letter has been underlined.

**8.37.** Let  $S$  be the set of  $k$ -element subsets of  $\mathbb{N}$ , weighted by the largest element in  $S$ . Find  $G_S$ . What happens if we weight a subset by its smallest element?

**8.38.** Fix  $k \in \mathbb{N}^+$ . Use the sum and product rules for weighted sets to find the generating function for the set of all compositions with  $k$  parts, weighted by the sum of the parts.

**8.39.** Compute the images of these partitions under Sylvester's bijection (see 8.22):

(a)  $(15, 5^2, 3^7, 1^9)$ ; (b)  $(7^5, 3^6, 1^3)$ ; (c)  $(11, 7, 5, 3)$ ; (d)  $(9^8)$ ; (e)  $(2n - 1, 2n - 3, \dots, 5, 3, 1)$ . (The notation  $5^2$  means two parts equal to 5, etc.)

**8.40.** Compute the images of these partitions under the inverse of Sylvester's bijection: (a)  $(15, 12, 10, 8, 6, 3, 1)$ ; (b)  $(28, 12, 7, 6, 2, 1)$ ; (c)  $(11, 7, 5, 3)$ ; (d)  $(21, 17, 16, 13, 11, 9, 8, 3, 2, 1)$ ; (e)  $(n, n - 1, \dots, 3, 2, 1)$ .

**8.41.** Compute the images of these partitions under Glaisher's bijection (see 8.23): (a)  $(9, 8, 5, 3, 1)$ ; (b)  $(28, 12, 7, 6, 2, 1)$ ; (c)  $(11, 7, 5, 3)$ ; (d)  $(21, 17, 16, 13, 11, 9, 8, 3, 2, 1)$ .

**8.42.** Compute the images of these partitions under the inverse of Glaisher's bijection: (a)  $(15, 5^2, 3^7, 1^9)$ ; (b)  $(13^2, 11^3, 7^4, 5, 3^2, 1^3)$ ; (c)  $(11, 7, 5, 3)$ ; (d)  $(9^8)$ ; (e)  $(1^n)$ .

**8.43.** Which partitions map to themselves under Glaisher's bijection? What about the generalized bijection in 8.24?

**8.44.** Let  $H$  and  $K$  be the maps in the proof of 8.24. (a) Find  $H(25, 17, 17, 10, 9, 6, 6, 5, 2, 2)$ , for  $d = 3, 4, 5$ . (b) Find  $K(8^{10}, 7^7, 2^{20}, 1^{30})$ , for  $d = 3, 5, 6$ .

**8.45.** Calculate the image of each partition under Franklin's involution (§8.7):

(a)  $(17, 16, 15, 14, 13, 10, 8, 7, 4)$ ; (b)  $(17, 16, 15, 14, 13, 10, 8)$ ; (c)  $(n, n - 1, \dots, 3, 2, 1)$ ; (d)  $(n)$ .



**8.46.** Find the generating function for the set of all integer partitions that satisfy each restriction below: (a) all parts are divisible by 3; (b) all parts are distinct and even; (c) odd parts appear at most twice; (d) each part is congruent to 1 or 4 mod 7; (e) for each  $i > 0$ , there are at most  $i$  parts of size  $i$ .

**8.47.** Give combinatorial interpretations for the coefficients in the following formal power series: (a)  $\prod_{i \geq 1} (1 - x^{5i})^{-1}$ ; (b)  $\prod_{i \geq 0} (1 + x^{6i+1})(1 + x^{6i+5})$ ; (c)  $\prod_{i \geq 2} (1 - x^{2i} + x^{4i})/(1 - x^i)$ .

**8.48.** (a) Show that the first Rogers-Ramanujan identity (see 8.25) can be written  $\prod_{n=0}^{\infty} \frac{1}{(1-x^{5n+1})(1-x^{5n+4})} = 1 + \sum_{k=1}^{\infty} \frac{x^{k^2}}{(1-x)(1-x^2)\cdots(1-x^k)}$ . (b) Find a similar formulation of the second Rogers-Ramanujan identity. (c) Verify the Rogers-Ramanujan identities for partitions of  $N = 12$  by explicitly listing all the partitions satisfying the relevant restrictions.

**8.49.** Give a detailed verification of the claim in 8.11 that the quadratic equation  $G^2 - G + x = 0$  has the unique solution  $G = (1 - \sqrt{1 - 4x})/2$  in  $\mathbb{Q}[[x]]$ .

**8.50.** Prove that a nonempty full binary tree with  $a$  leaves has  $a - 1$  non-leaf vertices.

**8.51.** Let  $f$  be the bijection in Figure 8.1. Compute  $f(T)$ , where  $T$  is the binary tree in Figure 2.12.

**8.52.** Let  $g$  be the bijection shown in Figure 8.2. Verify that the number of vertices in  $g(t)$  equals the number of leaves in  $t$ , for each full binary tree  $t$ .

**8.53.** (a) Describe the inverse of the bijection  $g$  shown in Figure 8.2. (b) Calculate the image of the ordered tree in Figure 3.18 under  $g^{-1}$ .

**8.54.** List all full binary trees with 5 leaves, and compute the image of each tree under the map  $g$  in Figure 8.2.

**8.55.** Verify that  $(R^n)_{n-1}/n = ((d/dx)^{n-1}R^n)_0/n!$  for all  $n \geq 1$  and  $R \in K[[x]]$ .

**8.56.** Give an algebraic proof of 8.24 using formal power series.

**8.57.** (a) Carefully verify that the maps  $H$  and  $K$  in 8.23 are two-sided inverses. (b) Repeat part (a) for the maps  $H$  and  $K$  in 8.24.

**8.58.** (a) Verify that the partition  $(2n, 2n-1, \dots, n+1)$  (one of the fixed points of Franklin's involution) has area  $n(3n+1)/2$ . (b) Verify that the partition  $(2n-1, 2n-2, \dots, n)$  has area  $n(3n-1)/2$ .

**8.59.** Carry out the computations showing how the equation  $C^* = \log(F)$  leads to formula (8.5).

**8.60.** Rewrite (8.5) as a sum over partitions of  $n$ .

**8.61.** Use (8.5) to compute the number of connected simple graphs with vertex set: (a)  $\{1, 2, 3, 4\}$ ; (b)  $\{1, 2, 3, 4, 5\}$ .

**8.62.** (a) Modify (8.5) to include a power of  $t$  that keeps track of the number of edges in the connected graph. (b) How many connected simple graphs with vertex set  $\{1, 2, 3, 4, 5, 6\}$  have exactly seven edges?

**8.63.** (a) Find the generating function for the set of all Dyck paths, where the weight of a path ending at  $(n, n)$  is  $n$ . (b) A *marked* Dyck path is a Dyck path in which one step (north or east) has been circled. Find the generating function for marked Dyck paths.

**8.64.** Recall that  $\sum_{n \geq 0} \sum_{k=0}^n \frac{S(n,k)}{n!} t^k x^n = \exp[t(e^x - 1)]$ . Use partial differentiation of this generating function to find generating functions for: (a) the set of set partitions, where one block in the partition has been circled; (b) the set of set partitions, where one element of one block in the partition has been circled.

**8.65.** (a) List all terms of length at most 5. (b) Use (a) and 8.14 to write down explicit formulas for the first five coefficients of the compositional inverse of  $x/R$  as combinations of the coefficients of  $R$ . (c) Use (b) to find the first five terms in the compositional inverse of  $x/(1 - 3x + 2x^2 + 5x^4)$ .

**8.66.** Use 8.15 to compute the compositional inverse of the following formal series: (a)  $xe^{2x}$ ; (b)  $x - x^2$ ; (c)  $x/(1 + ax)$ ; (d)  $x - 4x^4 + 4x^7$ .

**8.67.** Let  $S$  be the set of paths that start at  $(0,0)$  and take horizontal steps (right 1, up 0), vertical steps (right 0, up 1), and diagonal steps (right 1, up 1). By considering the final step of a path, find an equation satisfied by  $G_S$  and solve for  $G_S$ , taking the weight of a path ending at  $(c,d)$  to be: (a) the number of steps in the path; (b)  $c + d$ ; (c)  $c$ .

**8.68.** For fixed  $k \geq 1$ , find the generating function for integer partitions with: (a)  $k$  nonzero parts; (b)  $k$  nonzero distinct parts. (c) Deduce summation formulas for the infinite products  $\prod_{i \geq 1} (1 - x^i)^{-1}$  and  $\prod_{i \geq 1} (1 + x^i)$ .

**8.69.** A *ternary tree* is either  $\emptyset$  or a 4-tuple  $(\bullet, t_1, t_2, t_3)$ , where each  $t_i$  is itself a ternary tree. Find an equation satisfied by the generating function for ternary trees, weighted by number of vertices.

**8.70.** Let  $S$  be the set of ordered trees where every node has at most two children, weighted by the number of vertices. (a) Use the sum and product rules to find an equation satisfied by  $G_S$ . (b) Solve this equation for  $G_S$ . (c) How many trees in  $S$  have 7 vertices?

**8.71.** Find a formula for the number of simple digraphs with  $n$  vertices such that the graph obtained by erasing loops and ignoring the directions on the edges is connected.

**8.72.** Prove that the number of integer partitions of  $N$  in which no even part appears more than once equals the number of partitions of  $N$  in which no part appears 4 or more times.

**8.73.** Prove that the number of integer partitions of  $N$  that have no part equal to 1 and no parts that differ by 1 equals the number of partitions of  $N$  in which no part appears exactly once.

**8.74.** (a) Write down an infinite product that is the generating function for integer partitions with odd, distinct parts. (b) Show that the generating function for self-conjugate partitions (i.e., partitions such that  $\lambda' = \lambda$ ) is  $1 + \sum_{k=1}^{\infty} x^{k^2} / ((1 - x^2)(1 - x^4) \cdots (1 - x^{2k}))$ . (c) Find an area-preserving bijection between the sets of partitions in (a) and (b), and deduce an associated formal power series identity.

**8.75.** Evaluate  $\sum_{k=1}^{\infty} x^k (1 - x)^{-k}$ .

**8.76.** How many integer partitions of  $n$  have the form  $(i^j(i+1)^k)$  for some  $i, j, k > 0$ ?

**8.77. Dobinski's Formula.** Prove that the Bell numbers (see 2.51) satisfy  $B(n) = e^{-1} \sum_{k=0}^{\infty} (k^n / k!)$  for  $n \geq 0$ .

**8.78.** Show that, for all  $N \geq 0$ ,  $(-1)^N |\text{OddPar} \cap \text{DisPar} \cap \text{Par}(N)|$  equals  $|\{\mu \in \text{Par}(N) : \ell(\mu) \text{ is even}\}| - |\{\mu \in \text{Par}(N) : \ell(\mu) \text{ is odd}\}|$ .

**8.79.** (a) Use an involution on the set  $\text{Par} \times \text{DisPar}$  to give a combinatorial proof of the identity  $\prod_{n=1}^{\infty} \frac{1}{1-x^n} \prod_{n=1}^{\infty} (1-x^n) = 1$ . (b) More generally, for  $S \subseteq \mathbb{N}^+$ , prove combinatorially that  $\prod_{n \in S} \frac{1}{1-x^n} \prod_{n \in S} (1-x^n) = 1$ .

**8.80.** (a) Find a bijection from the set of terms of length  $n$  to the set of binary trees with  $n-1$  nodes. (b) Use (a) to formulate a version of 8.14 that expresses the coefficients in the compositional inverse of  $x/R$  as sums of suitably weighted binary trees.

**8.81.** (a) Find a bijection from the set of terms of length  $n$  to the set of Dyck paths ending at  $(n-1, n-1)$ . (b) Use (a) to formulate a version of 8.14 that expresses the coefficients in the compositional inverse of  $x/R$  as sums of suitably weighted Dyck paths.

**8.82.** Let  $d(n, k)$  be the number of derangements in  $S_n$  with  $k$  cycles. Find a formula for  $\sum_{n \geq 0} \sum_{k=0}^n \frac{d(n, k)}{n!} t^k x^n$ .

**8.83.** Compute  $\sum_{n \geq 0} d_n x^n / n!$ , where  $d_n$  is the number of derangements of  $n$  objects.

**8.84.** Let  $S_1(n, k)$  be the number of set partitions of  $\{1, 2, \dots, n\}$  into  $k$  blocks where no block consists of a single element. Find a formula for  $\sum_{n \geq 0} \sum_{k=0}^n \frac{S_1(n, k)}{n!} t^k x^n$ .

**8.85.** What is the generating function for set partitions in which all block sizes must belong to a given subset  $T \subseteq \mathbb{N}^+$ ?

**8.86.** Let  $A(n, k)$  be the number of ways to assign  $n$  people to  $k$  committees in such a way that each person belongs to exactly one committee, and each committee has one member designated as chairman. Find a formula for  $\sum_{n \geq 0} \sum_{k=0}^n \frac{A(n, k)}{n!} t^k x^n$ .

**8.87. Involution Proof of Euler's Partition Recursion.** (a) For fixed  $n \geq 1$ , prove  $\sum_{j \in \mathbb{Z}} (-1)^j p(n - (3j^2 + j)/2) = 0$  by verifying that the following map  $I$  is a sign-reversing involution with no fixed points. The domain of  $I$  is the set of pairs  $(j, \lambda)$  with  $j \in \mathbb{Z}$  and  $\lambda \in \text{Par}(n - (3j^2 + j)/2)$ . To define  $I(j, \lambda)$ , consider two cases. If  $\ell(\lambda) + 3j \geq \lambda_1$ , set  $I(j, \lambda) = (j-1, \mu)$  where  $\mu$  is formed by preceding the first part of  $\lambda$  by  $\ell(\lambda) + 3j$  and then decrementing all parts by 1. If  $\ell(\lambda) + 3j < \lambda_1$ , set  $I(j, \lambda) = (j+1, \nu)$  where  $\nu$  is formed by deleting the first part of  $\lambda$ , incrementing the remaining nonzero parts of  $\lambda$  by 1, and appending an additional  $\lambda_1 - 3j - \ell(\lambda) - 1$  parts equal to 1. (b) For  $n = 21$ ,  $j = 1$ ,  $\lambda = (5, 5, 4, 3, 2)$ , compute  $I(j, \lambda)$  and verify that  $I(I(j, \lambda)) = (j, \lambda)$ .

**8.88.** We say that an integer partition  $\lambda$  *extends* a partition  $\mu$  iff for all  $k$ ,  $k$  occurs in  $\lambda$  at least as often as  $k$  occurs in  $\mu$ ; otherwise,  $\lambda$  *avoids*  $\mu$ . Suppose  $\{\mu^i : i \geq 1\}$  and  $\{\nu^i : i \geq 1\}$  are two sequences of distinct, nonzero partitions such that for all finite  $S \subseteq \mathbb{N}^+$ ,  $\sum_{i \in S} |\mu^i| = \sum_{i \in S} |\nu^i|$ . (a) Prove that for every  $N$ , the number of partitions of  $N$  that avoid every  $\mu^i$  equals the number of partitions of  $N$  that avoid every  $\nu^i$ . (b) Show how 8.24 can be deduced from (a). (c) Use (a) to prove that the number of partitions of  $N$  into parts congruent to 1 or 5 mod 6 equals the number of partitions of  $N$  into distinct parts not divisible by 3.

## Notes

The applications of formal power series to combinatorial problems go well beyond the topics covered in this chapter. The texts [10, 127, 139] offer more detailed treatments of the uses of generating functions in combinatorics. Two more classical references are [90, 113]. For

an introduction to the vast subject of partition identities, the reader may consult [5, 102]. Sylvester's bijection appears in [129], Glaisher's bijection in [54], and Franklin's involution in [44]. The Rogers-Ramanujan identities are discussed in [106, 117]; Garsia and Milne gave the first bijective proof of these identities [49, 50]. Our treatment of compositional inversion closely follows the presentation in [107].

This page intentionally left blank

---

## Permutations and Group Actions

---

This chapter contains an introduction to some aspects of group theory that are directly related to combinatorial problems. The first part of the chapter gives the basic definitions of group theory and derives some fundamental properties of *symmetric groups*. We apply this material to give combinatorial derivations of the basic properties of determinants. The second part of the chapter discusses *group actions*, which have many applications to algebra and combinatorics. In particular, group actions can be used to solve counting problems in which symmetry must be taken into account. For example, how many ways can we color a  $5 \times 5$  chessboard with seven colors, if all rotations and reflections of a given colored board are considered the same? The theory of group actions provides systematic methods for solving problems like this one.

---

### 9.1 Definition and Examples of Groups

**9.1. Definition: Groups.** A *group* consists of a set  $G$  and a binary operation  $\star : G \times G \rightarrow G$  subject to the following axioms:

$$\begin{aligned} \forall x, y, z \in G, x \star (y \star z) &= (x \star y) \star z && \text{(associativity);} \\ \exists e \in G, \forall x \in G, x \star e &= x = e \star x && \text{(identity);} \\ \forall x \in G, \exists y \in G, x \star y &= e = y \star x && \text{(inverses).} \end{aligned}$$

The requirement that  $\star$  map  $G \times G$  into  $G$  is often stated explicitly as the following axiom:

$$\forall x, y \in G, x \star y \in G \quad \text{(closure).}$$

A group  $G$  is called *abelian* or *commutative* iff  $G$  satisfies the additional axiom

$$\forall x, y \in G, x \star y = y \star x \quad \text{(commutativity).}$$

**9.2. Example: Additive Groups.** The set  $\mathbb{Z}$  of all integers, with addition as the operation, is a commutative group. The identity element is  $e = 0$  and the (additive) inverse of  $x \in \mathbb{Z}$  is  $-x \in \mathbb{Z}$ . Similarly,  $\mathbb{Q}$  and  $\mathbb{R}$  and  $\mathbb{C}$  are all commutative groups under addition.  $\mathbb{N}^+$  is not a group under addition because there is no identity element *in the set*  $\mathbb{N}^+$ .  $\mathbb{N}$  is not a group under addition because  $1 \in \mathbb{N}$  has no additive inverse *in the set*  $\mathbb{N}$ . The three-element set  $S = \{-1, 0, 1\}$  is not a group under addition because closure fails ( $1 + 1 = 2 \notin S$ ).

**9.3. Example: Multiplicative Groups.** The set  $\mathbb{Q}^+$  of strictly positive rational numbers is a commutative group under multiplication. The identity element is  $e = 1$  and the inverse of  $a/b \in \mathbb{Q}^+$  is  $b/a$ . Similarly,  $\mathbb{R}^+$  is a group under multiplication. The set  $\mathbb{Q}$  is not a group under multiplication because  $0$  has no inverse. On the other hand,  $\mathbb{Q} \setminus \{0\}$ ,  $\mathbb{R} \setminus \{0\}$ , and  $\mathbb{C} \setminus \{0\}$  are groups under multiplication. So is the two-element set  $\{-1, 1\} \subseteq \mathbb{Q}$ , and the four-element set  $\{1, i, -1, -i\} \subseteq \mathbb{C}$ .

**9.4. Example: Symmetric Groups.** Let  $X$  be any set, and let  $\text{Sym}(X)$  be the set of all bijections  $f : X \rightarrow X$ . For  $f, g \in \text{Sym}(X)$ , define  $f \circ g$  to be the composite function that sends  $x \in X$  to  $f(g(x))$ . Then  $f \circ g \in \text{Sym}(X)$  since the composition of bijections is a bijection, so the axiom of closure holds. Given  $f, g, h \in \text{Sym}(X)$ , note that both of the functions  $(f \circ g) \circ h : X \rightarrow X$  and  $f \circ (g \circ h) : X \rightarrow X$  send  $x \in X$  to  $f(g(h(x)))$ . So these functions are equal, proving the axiom of associativity. Next, take  $e$  to be the bijection  $\text{id}_X : X \rightarrow X$ , which is defined by  $\text{id}_X(x) = x$  for all  $x \in X$ . One immediately checks that  $f \circ \text{id}_X = f = \text{id}_X \circ f$  for all  $f \in \text{Sym}(X)$ , so the identity axiom holds. Finally, given a bijection  $f \in \text{Sym}(X)$ , there exists an inverse function  $f^{-1} : X \rightarrow X$  that is also a bijection, and which satisfies  $f \circ f^{-1} = \text{id}_X = f^{-1} \circ f$ . So the axiom of inverses holds. This completes the verification that  $(\text{Sym}(X), \circ)$  is a group. This group is called the *symmetric group on  $X$* , and elements of  $\text{Sym}(X)$  are called *permutations of  $X$* . Symmetric groups play a central role in group theory and are closely related to group actions. In the special case when  $X = \{1, 2, \dots, n\}$ , we write  $S_n$  to denote the group  $\text{Sym}(X)$ .

Most of the groups  $\text{Sym}(X)$  are *not* commutative. For instance, consider  $f, g \in S_3$  given by

$$f(1) = 2, f(2) = 1, f(3) = 3; \quad g(1) = 3, g(2) = 2, g(3) = 1.$$

We see that  $(f \circ g)(1) = f(g(1)) = 3$ , whereas  $(g \circ f)(1) = g(f(1)) = 2$ . So  $f \circ g \neq g \circ f$ , and the axiom of commutativity fails.

**9.5. Example: Integers modulo  $n$ .** Let  $n$  be a fixed positive integer. Consider the set  $\mathbb{Z}_n = \underline{n} = \{0, 1, 2, \dots, n-1\}$ . We define a binary operation on  $\mathbb{Z}_n$  by setting, for all  $x, y \in \mathbb{Z}_n$ ,

$$x \oplus y = \begin{cases} x + y & \text{if } x + y < n; \\ x + y - n & \text{if } x + y \geq n. \end{cases}$$

Closure follows from this definition, once we note that  $0 \leq x + y \leq 2n - 2$  for  $x, y \in \mathbb{Z}_n$ . The identity element is 0. The inverse of 0 is 0, while for  $x > 0$  in  $\mathbb{Z}_n$ , the inverse of  $x$  is  $n - x \in \mathbb{Z}_n$ . To verify associativity, one may prove the relations

$$(x \oplus y) \oplus z = \begin{cases} x + y + z & \text{if } x + y + z < n; \\ x + y + z - n & \text{if } n \leq x + y + z < 2n; \\ x + y + z - 2n & \text{if } 2n \leq x + y + z < 3n; \end{cases} = x \oplus (y \oplus z), \quad (9.1)$$

which can be established by a tedious case analysis. Commutativity of  $\oplus$  follows from the definition and the commutativity of ordinary integer addition. We conclude that  $(\mathbb{Z}_n, \oplus)$  is a commutative group containing  $n$  elements. In particular, for every positive integer  $n$ , there exists a group of cardinality  $n$ .

**9.6. Definition: Multiplication Tables.** If  $(G, \star)$  is a group and  $G = \{x_1, \dots, x_n\}$  is finite, then a *multiplication table* for  $G$  is an  $n \times n$  table, with rows and columns labeled by the  $x_i$ 's, such that the element in row  $i$  and column  $j$  is  $x_i \star x_j$ . When the operation is written additively, we refer to this table as the *addition table* for  $G$ . It is customary, but not mandatory, to take  $x_1$  to be the identity element of  $G$ .

**9.7. Example.** The multiplication tables for  $(\{1, i, -1, -i\}, \times)$  and for  $(\mathbb{Z}_4, \oplus)$  are shown here:

$\times$	1	$i$	$-1$	$-i$	$\oplus$	0	1	2	3
1	1	$i$	$-1$	$-i$	0	0	1	2	3
$i$	$i$	$-1$	$-i$	1	1	1	2	3	0
$-1$	$-1$	$-i$	1	$i$	2	2	3	0	1
$-i$	$-i$	1	$i$	$-1$	3	3	0	1	2

The reader may notice a relationship between the two tables: each row within the table is

obtained from the preceding one by a cyclic shift one step to the left. (Using terminology to be discussed later, this happens because each of the two groups under consideration is cyclic of size four.)

One can define a group operation by specifying its multiplication table. For example, here is the table for another group of size four (which turns out not to be cyclic):

$\star$	$a$	$b$	$c$	$d$
$a$	$a$	$b$	$c$	$d$
$b$	$b$	$a$	$d$	$c$
$c$	$c$	$d$	$a$	$b$
$d$	$d$	$c$	$b$	$a$

The identity and inverse axioms can be checked from inspection of the table (here  $a$  is the identity, and every element is equal to its own inverse). There is no quick way to verify associativity by visual inspection of the table, but this axiom can be checked exhaustively using the table entries.

All of the groups in this example are commutative. This can be read off from the multiplication tables by noting the symmetry about the main diagonal line (the entry  $x_i \star x_j$  in row  $i$  and column  $j$  always equals the entry  $x_j \star x_i$  in row  $j$  and column  $i$ ).

## 9.2 Basic Properties of Groups

We now collect some facts about groups that follow from the defining axioms.

First, the identity element  $e$  in a group  $(G, \star)$  is *unique*. For, suppose  $e' \in G$  also satisfies the identity axiom. On one hand,  $e \star e' = e$  since  $e'$  is an identity element. On the other hand,  $e \star e' = e'$  since  $e$  is an identity element. So  $e = e'$ . (The very statement of the inverse axiom makes implicit use of the uniqueness of the identity.) We use the symbol  $e_G$  to denote the identity element of an abstract group  $G$ . When the operation is addition or multiplication, we write  $0_G$  or  $1_G$  instead, dropping the  $G$  if it is understood from context.

Similarly, the inverse of an element  $x$  in a group  $G$  is unique. For suppose  $y, y' \in G$  both satisfy the condition in the inverse axiom. Then

$$y = y \star e = y \star (x \star y') = (y \star x) \star y' = e \star y' = y'.$$

We denote the unique inverse of  $x$  in  $G$  by the symbol  $x^{-1}$ . When the operation is written additively, the symbol  $-x$  is used.

A product such as  $x \star y$  is often written  $xy$ , except in the additive case. The associativity axiom can be used to show that any parenthesization of a product  $x_1 x_2 \cdots x_n$  will give the same answer (see 2.148), so it is permissible to omit parentheses in products like these.

**9.8. Theorem: Cancellation Laws and Inverse Rules.** Suppose  $a, x, y$  are elements in a group  $G$ . (a)  $ax = ay$  implies  $x = y$  (left cancellation); (b)  $xa = ya$  implies  $x = y$  (right cancellation); (c)  $(x^{-1})^{-1} = x$ ; (d)  $(xy)^{-1} = y^{-1}x^{-1}$  (inverse rule for products).

*Proof.* Starting from  $ax = ay$ , multiply both sides on the left by  $a^{-1}$  to get  $a^{-1}(ax) = a^{-1}(ay)$ . Then the associativity axiom gives  $(a^{-1}a)x = (a^{-1}a)y$ ; the inverse axiom gives  $ex = ey$ ; and the identity axiom gives  $x = y$ . Right cancellation is proved similarly. Next, note that

$$(x^{-1})^{-1}x^{-1} = e = xx^{-1}$$



by the definition of the inverse of  $x$  and of  $x^{-1}$ ; right cancellation of  $x^{-1}$  yields  $(x^{-1})^{-1} = x$ . Similarly, routine calculations using the group axioms show that

$$(xy)^{-1}(xy) = e = (y^{-1}x^{-1})(xy),$$

so right cancellation of  $xy$  gives the inverse rule for products.  $\square$

**9.9. Definition: Exponent Notation.** Let  $G$  be a group written multiplicatively. Given  $x \in G$ , recursively define  $x^0 = 1 = e_G$  and  $x^{n+1} = x^n \star x$  for all  $n \geq 0$ . To define negative powers of  $x$ , set  $x^{-n} = (x^{-1})^n$  for all  $n > 0$ .

Informally, for positive  $n$ ,  $x^n$  is the product of  $n$  copies of  $x$ . For negative  $n$ ,  $x^n$  is the product of  $|n|$  copies of the inverse of  $x$ . Note in particular that  $x^1 = x$  and  $x^{-1}$  is the inverse of  $x$  (in accordance with the conventions introduced before this definition). When  $G$  is written additively, we write  $nx$  instead of  $x^n$ ; this denotes the sum of  $n$  copies of  $x$  for  $n > 0$ , or the sum of  $|n|$  copies of  $-x$  for  $n < 0$ .

**9.10. Theorem: Laws of Exponents.** Suppose  $G$  is a group written multiplicatively,  $x \in G$ , and  $m, n \in \mathbb{Z}$ . Then  $x^{m+n} = x^m x^n$  and  $x^{mn} = (x^n)^m$ . If  $x, y \in G$  satisfy  $xy = yx$ , then  $(xy)^n = x^n y^n$ . In additive notation, these results read:  $(m+n)x = mx + nx$ ;  $(mn)x = m(nx)$ ; and  $n(x+y) = nx + ny$  when  $x+y = y+x$ .

The idea of the proof is to use induction to establish the results for  $m, n \geq 0$ , and then use case analyses to handle the situations where  $m$  or  $n$  or both is negative. We leave the details as an exercise for the reader.

### 9.3 Notation for Permutations

Permutations and symmetric groups arise frequently in the theory of groups and group actions. So we will now develop some notation for describing permutations.

**9.11. Definition: Two-Line Form of a Function.** Let  $X$  be a finite set, and let  $x_1, \dots, x_n$  be a list of all the distinct elements of  $X$  in some fixed order. The *two-line form* of a function  $f : X \rightarrow X$  relative to this ordering is the array

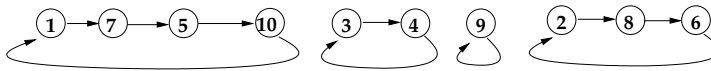
$$f = \left( \begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ f(x_1) & f(x_2) & \cdots & f(x_n) \end{array} \right).$$

If  $X = \{1, 2, \dots, n\}$ , we usually display the elements of  $X$  on the top line in the order  $1, 2, \dots, n$ .

**9.12. Example.** The notation  $f = \left( \begin{array}{ccccc} a & b & c & d & e \\ b & c & e & a & b \end{array} \right)$  defines a function on the set  $X = \{a, b, c, d, e\}$  such that  $f(a) = b$ ,  $f(b) = c$ ,  $f(c) = e$ ,  $f(d) = a$ , and  $f(e) = b$ . This function is not a permutation, since  $b$  occurs twice in the bottom row and  $d$  never occurs.

The notation  $g = \left( \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \end{array} \right)$  defines an element of  $S_5$  such that  $g(1) = 2$ ,  $g(2) = 4$ ,  $g(3) = 5$ ,  $g(4) = 1$ , and  $g(5) = 3$ . Observe that the *inverse* of  $g$  sends 2 to 1, 4 to 2, and so on. So, we obtain one possible two-line form of  $g^{-1}$  by interchanging the lines in the two-line form of  $g$ :

$$g^{-1} = \left( \begin{array}{ccccc} 2 & 4 & 5 & 1 & 3 \\ 1 & 2 & 3 & 4 & 5 \end{array} \right).$$



**FIGURE 9.1**

Digraph associated to the permutation  $h$ .

It is customary to write the numbers in the top line in increasing order. This can be accomplished by sorting the columns of the previous array:

$$g^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 5 & 2 & 3 \end{pmatrix}.$$

Recall that the group operation in  $\text{Sym}(X)$  is composition. We can compute the composition of two functions written in two-line form by tracing the effect of the composite function on each element. For instance,

$$\begin{pmatrix} a & b & c & d \\ b & d & a & c \end{pmatrix} \circ \begin{pmatrix} a & b & c & d \\ a & c & d & b \end{pmatrix} = \begin{pmatrix} a & b & c & d \\ b & a & c & d \end{pmatrix},$$

because the left side maps  $a$  to  $a$  and then to  $b$ ;  $b$  maps to  $c$  and then to  $a$ ; and so on.

If the ordering of  $X$  is fixed and known from context, we may omit the top line of the two-line form. This leads to one-line notation for a function defined on  $X$ .

**9.13. Definition: One-Line Form of a Function.** Let  $X = \{x_1 < x_2 < \cdots < x_n\}$  be a finite totally ordered set. The *one-line form* of a function  $f : X \rightarrow X$  is the array  $[f(x_1) \ f(x_2) \ \cdots \ f(x_n)]$ . We use square brackets to avoid a conflict with the cycle notation to be introduced below. Sometimes we omit the brackets, identifying  $f$  with the word  $f(x_1)f(x_2) \cdots f(x_n)$ .

**9.14. Example.** The functions  $f$  and  $g$  in the preceding example are given in one-line form by writing  $f = [b \ c \ e \ a \ b]$  and  $g = [2 \ 4 \ 5 \ 1 \ 3]$ . Note that the one-line form of an element of  $\text{Sym}(X)$  is literally a permutation of the elements of  $X$ , as defined in §1.4. This explains why elements of this group are called permutations.

**9.15. Cycle Notation for Permutations.** Assume  $X$  is a finite set. Recall from §3.6 that any function  $f : X \rightarrow X$  can be represented by a digraph with vertex set  $X$  and directed edges  $\{(i, f(i)) : i \in X\}$ . A digraph on  $X$  arises from a function in this way iff every vertex in  $X$  has outdegree 1. In 3.45 we proved that the digraph of a permutation is a disjoint union of directed cycles. For example, Figure 9.1 displays the digraph of the permutation

$$h = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 7 & 8 & 4 & 3 & 10 & 2 & 5 & 6 & 9 & 1 \end{pmatrix}.$$

We can describe a directed cycle in a digraph by traversing the edges in the cycle and listing the elements we encounter in the order of visitation, enclosing the whole list in parentheses. For example, the cycle containing 1 in Figure 9.1 can be described by writing  $(1, 7, 5, 10)$ . The cycle containing 9 is denoted by  $(9)$ . To describe the entire digraph of a permutation, we write down all the cycles in the digraph, one after the other. For example,  $h$  can be written in cycle notation as

$$h = (1, 7, 5, 10)(2, 8, 6)(3, 4)(9).$$

This cycle notation is not unique. We are free to begin our description of each cycle at any vertex in the cycle, and we are free to rearrange the order of the cycles. Furthermore, by convention it is permissible to omit some or all cycles of length 1. For example, some other cycle notations for  $h$  are

$$h = (5, 10, 1, 7)(3, 4)(9)(6, 2, 8) = (2, 8, 6)(4, 3)(7, 5, 10, 1).$$

To compute the inverse of a permutation written in cycle notation, we reverse the orientation of each cycle. For example,

$$h^{-1} = (10, 5, 7, 1)(6, 8, 2)(4, 3)(9).$$

**9.16. Example.** Using cycle notation, we can list the six elements of  $S_3$  as follows:

$$S_3 = \{(1)(2)(3), (1, 2), (1, 3), (2, 3), (1, 2, 3), (1, 3, 2)\}.$$

**9.17. Example.** To compose permutations written in cycle notation, we must see how the composite function acts on each element. For instance, consider the product  $(3, 5)(1, 2, 4) \circ (3, 5, 2, 1)$  in  $S_5$ . This composite function sends 1 to 3 and then 3 to 5, so 1 maps to 5. Next, 2 maps first to 1 and then to 2, so 2 maps to 2. Continuing similarly, we find that

$$(3, 5)(1, 2, 4) \circ (3, 5, 2, 1) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 3 & 1 & 4 \end{pmatrix} = (1, 5, 4)(2)(3).$$

With enough practice, one can proceed immediately to the cycle form of the answer without writing down the two-line form or other scratch work.

**9.18. Definition:  $k$ -cycles.** For  $k > 1$ , a permutation  $f \in \text{Sym}(X)$  whose digraph consists of one cycle of length  $k$  and all other cycles of length 1 is called a  $k$ -cycle.

**9.19. Remark.** We can view the cycle notation for a permutation  $f$  as a way of factorizing  $f$  in the group  $S_n$  into a product of cycles. For example,

$$(1, 7, 5, 10)(2, 8, 6)(3, 4)(9) = (1, 7, 5, 10) \circ (2, 8, 6) \circ (3, 4) \circ (9).$$

Here we have expressed the single permutation on the left side as a product of four other permutations in  $S_{10}$ . The stated equality may be verified by checking that both sides have the same effect on each  $k \in \{1, 2, \dots, 10\}$ .

**9.20. Definition:  $\text{cyc}(f)$  and  $\text{type}(f)$ .** Given a permutation  $f \in \text{Sym}(X)$ , let  $\text{cyc}(f)$  be the number of components (cycles) in the digraph for  $f$ . Let  $\text{type}(f)$  be the list of sizes of these components, including repetitions and written in weakly decreasing order.

Note that  $\text{type}(f)$  is an integer partition of  $n = |X|$ .

**9.21. Example.** The permutation  $h$  in Figure 9.1 has  $\text{cyc}(h) = 4$  and  $\text{type}(h) = (4, 3, 2, 1)$ . The identity element of  $S_n$ , namely  $\text{id} = (1)(2) \cdots (n)$ , has  $\text{cyc}(\text{id}) = n$  and  $\text{type}(\text{id}) = (1, \dots, 1)$ . Table 9.1 displays the 24 elements of  $S_4$  in cycle notation, collecting together all permutations with the same type and counting the number of permutations of each type. In 9.134, we will give a general formula for the number of permutations of  $n$  objects having a given type.

**TABLE 9.1**  
Elements of  $S_4$ .

Type	Permutations of this type	Count
(1, 1, 1, 1)	(1)(2)(3)(4)	1
(2, 1, 1)	(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)	6
(2, 2)	(1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)	3
(3, 1)	(1, 2, 3), (1, 2, 4), (1, 3, 4), (1, 3, 2), (1, 4, 2), (1, 4, 3), (2, 3, 4), (2, 4, 3)	8
(4)	(1, 2, 3, 4), (1, 2, 4, 3), (1, 3, 2, 4), (1, 3, 4, 2), (1, 4, 2, 3), (1, 4, 3, 2)	6

## 9.4 Inversions and Sign

In this section, we use inversions of permutations to define the *sign* function  $\text{sgn} : S_n \rightarrow \{+1, -1\}$ . We then study factorizations of permutations into products of transpositions to derive facts about the  $\text{sgn}$  function. Let us begin by recalling the definition of inversions (§6.2).

**9.22. Definition: Inversions and Sign of a Permutation.** Let  $w = w_1 w_2 \cdots w_n \in S_n$  be a permutation written in one-line form. An *inversion* of  $w$  is a pair of indices  $i < j$  such that  $w_i > w_j$ . The number of inversions of  $w$  is denoted  $\text{inv}(w)$ . Furthermore, the *sign* of  $w$  is defined to be  $\text{sgn}(w) = (-1)^{\text{inv}(w)}$ .

**9.23. Example.** Given  $w = 42531$ , we have  $\text{inv}(w) = 7$  and  $\text{sgn}(w) = -1$ . The seven inversions of  $w$  are (1, 2), (1, 4), (1, 5), (2, 5), (3, 4), (3, 5), and (4, 5). For instance, (1, 4) is an inversion because  $w_1 = 4 > 3 = w_4$ . The following table displays  $\text{inv}(f)$  and  $\text{sgn}(f)$  for all  $f \in S_3$ :

$f \in S_3$	$\text{inv}(f)$	$\text{sgn}(f)$
123	0	+1
132	1	-1
213	1	-1
231	2	+1
312	2	+1
321	3	-1

We want to understand how the group operation in  $S_n$  (composition of permutations) is related to inversions and sign. For this purpose, we introduce the concept of a *transposition*.

**9.24. Definition: Transpositions.** A *transposition* in  $S_n$  is a permutation  $f$  of the form  $(i, j)$ , for some  $i, j \leq n$ . Note that  $f(i) = j$ ,  $f(j) = i$ , and  $f(k) = k$  for all  $k \neq i, j$ . A *basic transposition* in  $S_n$  is a transposition  $(i, i + 1)$ , for some  $i < n$ .

The following lemmas illuminate the connection between basic transpositions and the process of sorting the one-line form of a permutation into increasing order.

**9.25. Lemma: Basic Transpositions and Sorting.** Let  $w = w_1 \cdots w_i w_{i+1} \cdots w_n \in S_n$  be a permutation in one-line form. For each  $i < n$ ,

$$w \circ (i, i + 1) = w_1 \cdots w_{i+1} w_i \cdots w_n.$$

So *right-multiplication by the basic transposition  $(i, i + 1)$  interchanges the elements in positions  $i$  and  $i + 1$  of  $w$ .*

*Proof.* Let us evaluate the function  $f = w \circ (i, i + 1)$  at each  $k \leq n$ . When  $k = i$ ,  $f(i) = w(i + 1)$ . When  $k = i + 1$ ,  $f(i + 1) = w(i)$ . When  $k \neq i$  and  $k \neq i + 1$ ,  $f(k) = k$ . So the one-line form of  $f$  is  $w_1 \cdots w_{i+1} w_i \cdots w_n$ , as desired.  $\square$

**9.26. Lemma: Basic Transpositions and Inversions.** Let  $w = w_1 \cdots w_n \in S_n$  be a permutation in one-line form, and let  $i < n$ . Then

$$\text{inv}(w \circ (i, i + 1)) = \begin{cases} \text{inv}(w) + 1 & \text{if } w_i < w_{i+1}; \\ \text{inv}(w) - 1 & \text{if } w_i > w_{i+1}. \end{cases}$$

Consequently, in all cases, we have

$$\text{sgn}(w \circ (i, i + 1)) = -\text{sgn}(w).$$

*Proof.* We use the result of the previous lemma to compare the inversions of  $w$  and  $w' = w \circ (i, i + 1)$ . Let  $j < k$  be two indices between 1 and  $n$ , and consider various cases. First, if  $j \neq i, i + 1$  and  $k \neq i, i + 1$ , then  $(j, k)$  is an inversion of  $w$  iff  $(j, k)$  is an inversion of  $w'$ , since  $w_j = w'_j$  and  $w_k = w'_k$ . Second, if  $j = i$  and  $k > i + 1$ , then  $(i, k)$  is an inversion of  $w$  iff  $(i + 1, k)$  is an inversion of  $w'$ , since  $w_i = w'_{i+1}$  and  $w_k = w'_k$ . Similar results hold in the cases  $(j = i + 1 < k)$ ,  $(j < k = i)$ , and  $(j < i, k = i + 1)$ . The critical case is when  $j = i$  and  $k = i + 1$ . If  $w_i < w_{i+1}$ , then  $(j, k)$  is an inversion of  $w'$  but not of  $w$ . If  $w_i > w_{i+1}$ , then  $(j, k)$  is an inversion of  $w$  but not of  $w'$ . This establishes the first formula in the lemma. The remaining formula follows since  $(-1)^{+1} = (-1)^{-1} = -1$ .  $\square$

The proof of the next lemma is left as an exercise.

**9.27. Lemma.** For all  $n \geq 1$ , the identity permutation  $\text{id} = 1, 2, \dots, n$  is the unique element of  $S_n$  satisfying  $\text{inv}(\text{id}) = 0$ . We have  $\text{sgn}(\text{id}) = +1$ .

If  $f = (i, i + 1)$  is a basic transposition, then the ordered pair  $(i, i + 1)$  is the only inversion of  $f$ , so  $\text{inv}(f) = 1$  and  $\text{sgn}(f) = -1$ . More generally, we now show that any transposition has sign  $-1$ .

**9.28. Lemma.** If  $f = (i, j)$  is any transposition, then  $\text{sgn}(f) = -1$ .

*Proof.* Since  $(i, j) = (j, i)$ , we may assume that  $i < j$ . Let us write  $f$  in two-line form:

$$f = \begin{pmatrix} 1 & \cdots & i & \cdots & j & \cdots & n \\ 1 & \cdots & j & \cdots & i & \cdots & n \end{pmatrix}.$$

We can find the inversions of  $f$  by inspecting the two-line form. The inversions are: all  $(i, k)$  with  $i < k \leq j$ ; and all  $(k, j)$  with  $i < k < j$ . There are  $j - i$  inversions of the first type and  $j - i - 1$  inversions of the second type, hence  $2(j - i) - 1$  inversions total. Since this number is odd, we conclude that  $\text{sgn}(f) = -1$ .  $\square$

**9.29. Theorem: Inversions and Sorting.** Let  $w = w_1 w_2 \cdots w_n \in S_n$  be a permutation in one-line form. The number  $\text{inv}(w)$  is the minimum number of steps required to sort the word  $w$  into increasing order by repeatedly interchanging two adjacent elements. Furthermore,  $w$  can be factored in  $S_n$  into the product of  $\text{inv}(w)$  basic transpositions.

*Proof.* Given  $w \in S_n$ , it is certainly possible to sort  $w$  into increasing order in finitely many steps by repeatedly swapping adjacent elements. For instance, we can move 1 to the far left position in at most  $n - 1$  moves, then move 2 to its proper position in at most  $n - 2$  moves, and so on. Let  $m$  be the *minimum* number of moves of this kind that are needed to sort  $w$ . By 9.25, we can accomplish each sorting move by starting with  $w$  and repeatedly multiplying on the right by suitable basic transpositions. Each such multiplication either increases or decreases the inversion count by 1, according to 9.26. At the end, we have transformed  $w$  into the identity permutation. Combining these observations, we see that  $0 = \text{inv}(\text{id}) \geq \text{inv}(w) - m$ , so that  $m \geq \text{inv}(w)$ . On the other hand, consider the following particular sequence of sorting moves starting from  $w$ . If the current permutation  $w^*$  is not the identity, there exists a smallest index  $i$  with  $w_i^* > w_{i+1}^*$ . Apply the basic transposition  $(i, i + 1)$ , which reduces  $\text{inv}(w^*)$  by 1, and continue. This sorting method will end in exactly  $\text{inv}(w)$  steps, since  $\text{id}$  is the unique permutation with zero inversions. This proves it is possible to sort  $w$  in  $\text{inv}(w)$  steps, so that  $m \leq \text{inv}(w)$ .

To prove the last part of the theorem, recall that the sorting process just described can be implemented by right-multiplying by suitable basic transpositions. We therefore have an equation in  $S_n$  of the form

$$w \circ (i_1, i_1 + 1) \circ (i_2, i_2 + 1) \circ \cdots \circ (i_m, i_m + 1) = \text{id}.$$

Solving for  $w$ , and using the fact that  $(i, j)^{-1} = (j, i) = (i, j)$ , we get

$$w = (i_m, i_m + 1) \circ \cdots \circ (i_2, i_2 + 1) \circ (i_1, i_1 + 1),$$

which expresses  $w$  as a product of  $m$  basic transpositions. □

**9.30. Example.** Let us apply the sorting algorithm in the preceding proof to write  $w = 42531$  as a product of  $\text{inv}(w) = 7$  basic transpositions. Since  $4 > 2$ , we first multiply  $w$  on the right by  $(1, 2)$  to obtain

$$w \circ (1, 2) = 24531.$$

(Observe that  $\text{inv}(24531) = 6 = \text{inv}(w) - 1$ .) Next, since  $5 > 3$ , we multiply on the right by  $(3, 4)$  to get

$$w \circ (1, 2) \circ (3, 4) = 24351.$$

The computation continues as follows:

$$\begin{aligned} w \circ (1, 2) \circ (3, 4) \circ (2, 3) &= 23451; \\ w \circ (1, 2) \circ (3, 4) \circ (2, 3) \circ (4, 5) &= 23415; \\ w \circ (1, 2) \circ (3, 4) \circ (2, 3) \circ (4, 5) \circ (3, 4) &= 23145; \\ w \circ (1, 2) \circ (3, 4) \circ (2, 3) \circ (4, 5) \circ (3, 4) \circ (2, 3) &= 21345; \\ w \circ (1, 2) \circ (3, 4) \circ (2, 3) \circ (4, 5) \circ (3, 4) \circ (2, 3) \circ (1, 2) &= 12345 = \text{id}. \end{aligned}$$

We now solve for  $w$ , which has the effect of reversing the order of the basic transpositions we used to reach the identity:

$$w = (1, 2) \circ (2, 3) \circ (3, 4) \circ (4, 5) \circ (2, 3) \circ (3, 4) \circ (1, 2).$$

It is also possible to find such a factorization by starting with the identity word and “un-sorting” to reach  $w$ . Here it will not be necessary to reverse the order of the transpositions

at the end. We illustrate this idea with the following computation:

$$\begin{aligned}
 \text{id} &= 12345; \\
 \text{id} \circ (3, 4) &= 12435; \\
 \text{id} \circ (3, 4) \circ (2, 3) &= 14235; \\
 \text{id} \circ (3, 4) \circ (2, 3) \circ (1, 2) &= 41235; \\
 \text{id} \circ (3, 4) \circ (2, 3) \circ (1, 2) \circ (2, 3) &= 42135; \\
 \text{id} \circ (3, 4) \circ (2, 3) \circ (1, 2) \circ (2, 3) \circ (4, 5) &= 42153; \\
 \text{id} \circ (3, 4) \circ (2, 3) \circ (1, 2) \circ (2, 3) \circ (4, 5) \circ (3, 4) &= 42513; \\
 \text{id} \circ (3, 4) \circ (2, 3) \circ (1, 2) \circ (2, 3) \circ (4, 5) \circ (3, 4) \circ (4, 5) &= 42531 = w.
 \end{aligned}$$

So  $w = (3, 4) \circ (2, 3) \circ (1, 2) \circ (2, 3) \circ (4, 5) \circ (3, 4) \circ (4, 5)$ . Observe that this is a different factorization of  $w$  from the one obtained earlier, although both involve seven basic transpositions. This shows that *factorizations of permutations into products of basic transpositions are not unique*. It is also possible to find factorizations involving more than seven factors, by interchanging two entries that are already in the correct order during the sorting of  $w$  into  $\text{id}$ . So the number of factors in such factorizations is not unique either; but we will see shortly that the *parity* of the number of factors (odd or even) *is* uniquely determined by  $w$ . In fact, the parity is odd when  $\text{sgn}(w) = -1$  and even when  $\text{sgn}(w) = +1$ .

We now have enough machinery to prove the fundamental properties of  $\text{sgn}$ .

**9.31. Theorem: Properties of Sign.** (a) For all  $f, g \in S_n$ ,  $\text{sgn}(f \circ g) = \text{sgn}(f) \cdot \text{sgn}(g)$ .  
 (b) For all  $f \in S_n$ ,  $\text{sgn}(f^{-1}) = \text{sgn}(f)$ .

*Proof.* (a) If  $g = \text{id}$ , then the result is true since  $f \circ g = f$  and  $\text{sgn}(g) = 1$  in this case. If  $t = (i, i+1)$  is a basic transposition, then 9.26 shows that  $\text{sgn}(f \circ t) = -\text{sgn}(f)$ . Given a non-identity permutation  $g$ , use 9.29 to write  $g$  as a nonempty product of basic transpositions, say  $g = t_1 \circ t_2 \circ \cdots \circ t_k$ . Then, for every  $f \in S_n$ , iteration of 9.26 gives

$$\begin{aligned}
 \text{sgn}(f \circ g) &= \text{sgn}(ft_1 \cdots t_{k-1}t_k) = -\text{sgn}(ft_1 \cdots t_{k-1}) \\
 &= (-1)^2 \text{sgn}(ft_1 \cdots t_{k-2}) = \cdots = (-1)^k \text{sgn}(f).
 \end{aligned}$$

In particular, this equation is true when  $f = \text{id}$ ; in that case, we obtain  $\text{sgn}(g) = (-1)^k$ . Using this fact in the preceding equation produces  $\text{sgn}(f \circ g) = \text{sgn}(g) \text{sgn}(f) = \text{sgn}(f) \text{sgn}(g)$  for all  $f \in S_n$ .

(b) By part (a),  $\text{sgn}(f) \cdot \text{sgn}(f^{-1}) = \text{sgn}(f \circ f^{-1}) = \text{sgn}(\text{id}) = +1$ . If  $\text{sgn}(f) = +1$ , it follows that  $\text{sgn}(f^{-1}) = +1$ . If instead  $\text{sgn}(f) = -1$ , then it follows that  $\text{sgn}(f^{-1}) = -1$ .  $\square$

Iteration of 9.31 shows that

$$\text{sgn}(f_1 \circ \cdots \circ f_k) = \prod_{i=1}^k \text{sgn}(f_i). \quad (9.2)$$

**9.32. Theorem: Factorizations into Transpositions.** Let  $f = t_1 \circ t_2 \circ \cdots \circ t_k$  be *any* factorization of  $f \in S_n$  into a product of transpositions (not necessarily basic ones). Then  $\text{sgn}(f) = (-1)^k$ . In particular, the parity of  $k$  (odd or even) is uniquely determined by  $f$ .

*Proof.* By 9.28,  $\text{sgn}(t_i) = -1$  for all  $i$ . The conclusion now follows by setting  $f_i = t_i$  in (9.2).  $\square$

**9.33. Theorem: Sign of a  $k$ -cycle.** The sign of any  $k$ -cycle  $(i_1, i_2, \dots, i_k)$  is  $(-1)^{k-1}$ .

*Proof.* The result is already known for  $k = 1$  and  $k = 2$ . For  $k > 2$ , one may check that the given  $k$ -cycle can be written as the following product of  $k - 1$  transpositions:

$$(i_1, i_2, \dots, i_k) = (i_1, i_2) \circ (i_2, i_3) \circ (i_3, i_4) \circ \cdots \circ (i_{k-1}, i_k).$$

So the result follows from 9.32.  $\square$

We can now show that the sign of a permutation  $f$  is completely determined by  $\text{type}(f)$ .

**9.34. Theorem: Cycle Type and Sign.** Suppose  $f \in S_n$  has  $\text{type}(f) = \mu$ . Then

$$\text{sgn}(f) = \prod_{i=1}^{\ell(\mu)} (-1)^{\mu_i - 1} = (-1)^{n - \ell(\mu)} = (-1)^{n - \text{cyc}(f)}.$$

*Proof.* Let the cycle decomposition of  $f$  be  $f = C_1 \circ \cdots \circ C_{\ell(\mu)}$ , where  $C_i$  is a  $\mu_i$ -cycle. The result follows from the relations  $\text{sgn}(f) = \prod_{i=1}^{\ell(\mu)} \text{sgn}(C_i)$  and  $\text{sgn}(C_i) = (-1)^{\mu_i - 1}$ .  $\square$

**9.35. Example.** The permutation  $f = (4, 6, 2, 8)(3, 9, 1)(5, 10, 7)$  in  $S_{10}$  has  $\text{sgn}(f) = (-1)^{10-3} = -1$ .

## 9.5 Determinants

In the next three sections, we interrupt our exposition of group theory to give an application of the preceding material to determinants. We will see that the combinatorial properties of permutations underlie many commonly used facts about determinants.

**9.36. Definition: Matrix Rings.** For every commutative ring  $R$  and positive integer  $n$ , let  $M_n(R)$  be the set of  $n \times n$  matrices with entries in  $R$ . Formally, an element of  $M_n(R)$  is a function  $A : \{1, 2, \dots, n\} \times \{1, 2, \dots, n\} \rightarrow R$ .  $A(i, j)$  is called the  $i, j$ -entry of  $A$ . We often display  $A$  as a square array in which  $A(i, j)$  appears in row  $i$  and column  $j$ . For  $A, B \in M_n(R)$  and  $c \in R$ , define  $A + B$ ,  $AB$ , and  $cA$  by setting

$$\begin{aligned} (A + B)(i, j) &= A(i, j) + B(i, j); \\ (AB)(i, j) &= \sum_{k=1}^n A(i, k)B(k, j); \\ (cA)(i, j) &= c(A(i, j)) \quad (1 \leq i, j \leq n). \end{aligned}$$

Routine verifications (see 2.151) show that  $M_n(R)$  with these operations is a ring, whose multiplicative identity element  $I_n$  is given by  $I_n(i, j) = 1_R$  if  $i = j$ , and  $I_n(i, j) = 0_R$  if  $i \neq j$ . One also checks that  $M_n(R)$  is non-commutative if  $n > 1$  and  $R \neq \{0\}$ .

**9.37. Definition: Determinants.** For a matrix  $A \in M_n(R)$ , the *determinant* of  $A$  is

$$\det(A) = \sum_{w \in S_n} \text{sgn}(w) \prod_{i=1}^n A(i, w(i)) \in R.$$

**9.38. Example.** When  $n = 1$ ,  $\det(A) = A(1, 1)$ . When  $n = 2$ , the possible permutations  $w$



(in one-line form) are  $w = 12$  with  $\text{sgn}(w) = +1$ , and  $w = 21$  with  $\text{sgn}(w) = -1$ . Therefore, the definition gives

$$\det(A) = \det \begin{bmatrix} A(1,1) & A(1,2) \\ A(2,1) & A(2,2) \end{bmatrix} = +A(1,1)A(2,2) - A(1,2)A(2,1).$$

When  $n = 3$ , the definition and the table in 9.23 lead to the formula

$$\begin{aligned} \det(A) &= \det \begin{bmatrix} A(1,1) & A(1,2) & A(1,3) \\ A(2,1) & A(2,2) & A(2,3) \\ A(3,1) & A(3,2) & A(3,3) \end{bmatrix} \\ &= +A(1,1)A(2,2)A(3,3) - A(1,1)A(2,3)A(3,2) - A(1,2)A(2,1)A(3,3) \\ &\quad + A(1,2)A(2,3)A(3,1) + A(1,3)A(2,1)A(3,2) - A(1,3)A(2,2)A(3,1). \end{aligned}$$

In general, we see that  $\det(A)$  is a sum of  $n!$  signed terms. A given term arises by choosing one factor  $A(i, w(i))$  from each row of  $A$ ; since  $w$  is a permutation, each of the chosen factors must come from a different column of  $A$ . The term in question is the product of the  $n$  chosen factors, times  $\text{sgn}(w)$ . Since  $\text{sgn}(w) = (-1)^{\text{inv}(w)}$ , the sign attached to this term depends on the parity of the number of basic transpositions needed to sort the column indices  $w(1), w(2), \dots, w(n)$  into increasing order.

The next result shows that we can replace  $A(i, w(i))$  by  $A(w(i), i)$  in the defining formula for  $\det(A)$ . This corresponds to interchanging the roles of rows and columns in the description above.

**9.39. Definition: Transpose of a Matrix.** Given  $A \in M_n(R)$ , the *transpose* of  $A$  is the matrix  $A^t \in M_n(R)$  such that  $A^t(i, j) = A(j, i)$  for all  $i, j \leq n$ .

**9.40. Theorem: Determinant of a Transpose.** For all  $A \in M_n(R)$ ,  $\det(A^t) = \det(A)$ .

*Proof.* By definition,

$$\det(A^t) = \sum_{w \in S_n} \text{sgn}(w) \prod_{k=1}^n A^t(k, w(k)) = \sum_{w \in S_n} \text{sgn}(w) \prod_{k=1}^n A(w(k), k).$$

For a fixed  $w \in S_n$ , we make a change of variables in the product indexed by  $w$  by letting  $j = w(k)$ , so  $k = w^{-1}(j)$ . Since  $w$  is a permutation and  $R$  is commutative, we have

$$\prod_{k=1}^n A(w(k), k) = \prod_{j=1}^n A(j, w^{-1}(j))$$

because the second product contains the same factors as the first product in a different order (cf. 2.149). We now calculate

$$\det(A^t) = \sum_{w \in S_n} \text{sgn}(w) \prod_{j=1}^n A(j, w^{-1}(j)) = \sum_{w \in S_n} \text{sgn}(w^{-1}) \prod_{j=1}^n A(j, w^{-1}(j)).$$

Now consider the change of variable  $v = w^{-1}$ . As  $w$  ranges over  $S_n$ , so does  $v$ , since  $w \mapsto w^{-1}$  is a bijection on  $S_n$ . Furthermore, we can reorder the terms of the sum since addition in  $R$  is commutative (see 2.149). We conclude that

$$\det(A^t) = \sum_{v \in S_n} \text{sgn}(v) \prod_{j=1}^n A(j, v(j)) = \det(A). \quad \square$$

Next we derive a formula for the determinant of an upper-triangular matrix.

**9.41. Theorem: Determinant of Triangular and Diagonal Matrices.** Suppose  $A \in M_n(R)$  satisfies  $A(i, j) = 0$  whenever  $i > j$ . Then  $\det(A) = \prod_{i=1}^n A(i, i)$ . Consequently, if  $A$  is either upper-triangular, lower-triangular, or diagonal, then  $\det(A)$  is the product of the diagonal entries of  $A$ .

*Proof.* By definition,  $\det(A) = \sum_{w \in S_n} \text{sgn}(w) \prod_{i=1}^n A(i, w(i))$ . In order for a given summand to be nonzero, we must have  $i \leq w(i)$  for all  $i \leq n$ . Since  $w$  is a permutation, we successively deduce that  $w(n) = n, w(n-1) = n-1, \dots, w(1) = 1$ . Thus, the only possibly nonzero summand comes from  $w = \text{id}$ . Since  $\text{sgn}(\text{id}) = +1$  and  $\text{id}(i) = i$  for all  $i$ , the stated formula for  $\det(A)$  follows when  $A$  is upper-triangular. The result for lower-triangular  $A$  follows by considering  $A^t$ . Since diagonal matrices are upper-triangular, the proof is complete.  $\square$

**9.42. Corollary: Determinant of Identity Matrix.** For all  $n \in \mathbb{N}^+$ ,  $\det(I_n) = 1_R$ .

## 9.6 Multilinearity and Laplace Expansions

This section continues our development of the properties of determinants.

**9.43. Definition:  $R$ -Linear Maps.** Let  $R$  be a commutative ring and  $n \in \mathbb{N}^+$ . A map  $T : R^n \rightarrow R$  is called  $R$ -linear iff  $T(v+z) = T(v) + T(z)$  and  $T(cv) = cT(v)$  for all  $v, z \in R^n$  and all  $c \in R$ .

**9.44. Example.** Suppose  $b_1, \dots, b_n \in R$  are fixed constants, and  $T : R^n \rightarrow R$  is defined by

$$T(v_1, \dots, v_n) = b_1 v_1 + b_2 v_2 + \dots + b_n v_n.$$

It is routine to check that the map  $T$  is  $R$ -linear. Conversely, one can show that every  $R$ -linear map from  $R^n$  to  $R$  must be of this form.

**9.45. Theorem: Multilinearity of Determinants.** Let  $A \in M_n(R)$ , and let  $k \leq n$  be a fixed row index. For every row vector  $v \in R^n$ , let  $A[v]$  denote the matrix  $A$  with row  $k$  replaced by  $v$ . Then the map  $T : R^n \rightarrow R$  given by  $T(v) = \det(A[v])$  is  $R$ -linear. A similar result holds for the columns of  $A$ .

*Proof.* By 9.44, it suffices to show that there exist constants  $b_1, \dots, b_n \in R$  such that for all  $v = (v_1, v_2, \dots, v_n) \in R^n$ ,

$$T(v) = b_1 v_1 + b_2 v_2 + \dots + b_n v_n. \quad (9.3)$$

To establish this, consider the defining formula for  $\det(A[v])$ :

$$T(v) = \det(A[v]) = \sum_{w \in S_n} \text{sgn}(w) \prod_{i=1}^n A[v](i, w(i)) = \sum_{w \in S_n} \text{sgn}(w) \left[ \prod_{\substack{i=1 \\ i \neq k}}^n A(i, w(i)) \right] v_{w(k)}.$$

The terms in brackets depend only on the fixed matrix  $A$ , not on  $v$ . So (9.3) holds with

$$b_j = \sum_{\substack{w \in S_n \\ w(k)=j}} \text{sgn}(w) \prod_{\substack{i=1 \\ i \neq k}}^n A(i, w(i)) \quad (1 \leq j \leq n). \quad (9.4)$$

To obtain the multilinearity result for the columns of  $A$ , apply the result just proved to  $A^t$ .  $\square$

We sometimes use the following notation when invoking the multilinearity of determinants. For  $A \in M_n(R)$ , let  $A_1, A_2, \dots, A_n$  denote the  $n$  rows of  $A$ ; thus each  $A_i$  lies in  $R^n$ . We write  $\det(A) = \det(A_1, \dots, A_n)$ , viewing the determinant as a function of  $n$  arguments (row vectors). The previous result says that if we fix any  $n-1$  of these arguments and let the other one vary, the resulting map  $v \mapsto \det(A_1, \dots, v, \dots, A_n)$  (for  $v \in R^n$ ) is  $R$ -linear.

**9.46. Theorem: Alternating Property of Determinants.** If  $A \in M_n(R)$  has two equal rows or two equal columns, then  $\det(A) = 0$ .

*Proof.* Recall  $\det(A)$  is a sum of  $n!$  signed terms of the form  $T(w) = \operatorname{sgn}(w) \prod_{i=1}^n A(i, w(i))$ , where  $w$  ranges over  $S_n$ . Suppose rows  $r$  and  $s$  of  $A$  are equal, so  $A(r, k) = A(s, k)$  for all  $k$ . We will define an involution  $I$  on  $S_n$  with no fixed points such that  $T(I(w)) = -T(w)$  for all  $w \in S_n$ . It will follow that the  $n!$  terms cancel in pairs, so that  $\det(A) = 0$ . Define  $I(w) = w \circ (r, s)$  for  $w \in S_n$ ; evidently  $I \circ I = \operatorname{id}_{S_n}$  and  $I$  has no fixed points. On one hand,  $\operatorname{sgn}(I(w)) = \operatorname{sgn}(w) \cdot \operatorname{sgn}((r, s)) = -\operatorname{sgn}(w)$  by 9.31. On the other hand,

$$\begin{aligned} \prod_{i=1}^n A(i, [w \circ (r, s)](i)) &= A(r, w(s))A(s, w(r)) \prod_{i \neq r, s} A(i, w(i)) \\ &= A(r, w(r))A(s, w(s)) \prod_{i \neq r, s} A(i, w(i)) = \prod_{i=1}^n A(i, w(i)). \end{aligned}$$

Combining these facts, we see that  $T(I(w)) = -T(w)$ , as desired. If  $A$  has two equal columns, then  $A^t$  has two equal rows, so  $\det(A) = \det(A^t) = 0$ .  $\square$

**9.47. Theorem: Effect of Elementary Row Operations on Determinants.** Let  $A \in M_n(R)$ , let  $j, k$  be distinct indices, and let  $c \in R$ .

- (a) If  $B$  is obtained from  $A$  by multiplying row  $j$  by  $c$ , then  $\det(B) = c \det(A)$ .
- (b) If  $B$  is obtained from  $A$  by interchanging rows  $j$  and  $k$ , then  $\det(B) = -\det(A)$ .
- (c) If  $B$  is obtained from  $A$  by adding  $c$  times row  $j$  to row  $k$ , then  $\det(B) = \det(A)$ .

Analogous results hold for elementary column operations.

*Proof.* Part (a) is a special case of the multilinearity of determinants (see 9.45). Part (b) is a consequence of multilinearity and the alternating property. Specifically, define  $T : R^n \times R^n \rightarrow R$  by letting  $T(v, w) = \det(A_1, \dots, v, \dots, w, \dots, A_n)$  (where the  $v$  and  $w$  occur in positions  $j$  and  $k$ ). Since  $\det$  is multilinear and alternating, we get

$$0 = T(v+w, v+w) = T(v, v) + T(w, v) + T(v, w) + T(w, w) = T(w, v) + T(v, w) \quad (v, w \in R^n).$$

Thus,  $T(w, v) = -T(v, w)$  for all  $v, w$ , which translates to statement (b) after taking  $v = A_j$  and  $w = A_k$ . Part (c) follows for similar reasons, since

$$T(v, cv + w) = cT(v, v) + T(v, w) = T(v, w). \quad \square$$

**9.48. Theorem: Laplace Expansions of Determinants.** For  $A \in M_n(R)$  and  $i, j \leq n$ , let  $A[i|j]$  be the matrix in  $M_{n-1}(R)$  obtained by deleting row  $i$  and column  $j$  of  $A$ . For  $1 \leq k \leq n$ , we have

$$\begin{aligned} \det(A) &= \sum_{i=1}^n (-1)^{i+k} A(i, k) \det(A[i|k]) \quad (\text{expansion along column } k) \\ &= \sum_{j=1}^n (-1)^{j+k} A(k, j) \det(A[k|j]) \quad (\text{expansion along row } k). \end{aligned}$$

*Proof.* Let us first prove the Laplace expansion formula along row  $k = n$ . By the proof of multilinearity (see equations (9.3) and (9.4)), we know that

$$\det(A) = b_1 A(n, 1) + b_2 A(n, 2) + \cdots + b_n A(n, n)$$

where

$$b_j = \sum_{\substack{w \in S_n \\ w(n)=j}} \operatorname{sgn}(w) \prod_{i=1}^{n-1} A(i, w(i)) \quad (1 \leq j \leq n).$$

Comparing to the desired formula, we need only show that  $b_j = (-1)^{j+n} \det(A[n|j])$  for all  $j$ .

Fix an index  $j$ . Let  $S_{n,j} = \{w \in S_n : w(n) = j\}$ . We define a bijection  $f : S_{n,j} \rightarrow S_{n-1}$  as follows. Every  $w \in S_{n,j}$  can be written in one-line form as  $w = w_1 w_2 \cdots w_{n-1} w_n$  where  $w_n = j$ . Define  $f(w) = w'_1 w'_2 \cdots w'_{n-1}$  where  $w'_t = w_t$  if  $w_t < j$ , and  $w'_t = w_t - 1$  if  $w_t > j$ . In other words, we drop the  $j$  at the end of  $w$  and decrement all letters larger than  $j$ . The inverse map increments all letters  $\geq j$  and then adds a  $j$  at the end. Observe that the deletion of  $j$  decreases  $\operatorname{inv}(w)$  by  $n - j$  (the number of letters to the left of  $j$  that are greater than  $j$ ), and the decrementing operation has no further effect on the inversion count. So,  $\operatorname{inv}(f(w)) = \operatorname{inv}(w) - (n - j)$  and  $\operatorname{sgn}(f(w)) = (-1)^{j+n} \operatorname{sgn}(w)$ . We also note that for  $w' = f(w)$ , we have  $A(i, w(i)) = A[n|j](i, w'(i))$  for all  $i < n$ , since all columns in  $A$  after column  $j$  get shifted one column left when column  $j$  is deleted. Now use the bijection  $f$  to change the summation variable in the formula for  $b_j$ . Writing  $w' = f(w)$ , we obtain

$$\begin{aligned} b_j &= \sum_{w \in S_{n,j}} \operatorname{sgn}(w) \prod_{i=1}^{n-1} A(i, w(i)) \\ &= \sum_{w' \in S_{n-1}} (-1)^{j+n} \operatorname{sgn}(w') \prod_{i=1}^{n-1} A[n|j](i, w'(i)) = (-1)^{j+n} \det(A[n|j]). \end{aligned}$$

The Laplace expansion along an arbitrary row  $k$  follows from the special case  $k = n$ . Given  $k$ , let  $B$  be the matrix obtained from  $A$  by successively interchanging row  $k$  with row  $k + 1$ ,  $k + 2$ ,  $\dots$ ,  $n$ . These  $n - k$  row interchanges multiply the determinant by  $(-1)^{n-k}$ . It is evident that  $B(n, j) = A(k, j)$  and  $B[n|j] = A[k|j]$  for all  $j$ . So

$$\begin{aligned} \det(A) &= (-1)^{n-k} \det(B) = (-1)^{k-n} \sum_{j=1}^n (-1)^{j+n} B(n, j) \det(B[n|j]) \\ &= \sum_{j=1}^n (-1)^{j+k} A(k, j) \det(A[k|j]). \end{aligned}$$

Finally, to derive the Laplace expansion along column  $k$ , pass to transposes:

$$\begin{aligned} \det(A) &= \det(A^t) = \sum_{j=1}^n (-1)^{j+k} A^t(k, j) \det(A^t[k|j]) \\ &= \sum_{j=1}^n (-1)^{j+k} A(j, k) \det(A[j|k]). \quad \square \end{aligned}$$

We now use Laplace expansions to derive the classical formula for the inverse of a matrix.

**9.49. Definition: Classical Adjoint of a Matrix.** Given  $A \in M_n(R)$ , let  $\operatorname{adj} A \in M_n(R)$  be the matrix with  $i, j$ -entry  $(-1)^{i+j} \det(A[j|i])$  for  $i, j \leq n$ .

The next result explains why we wrote  $A[j|i]$  instead of  $A[i|j]$  in the preceding definition.

**9.50. Theorem: Adjoint Formula.** For all  $A \in M_n(R)$ , we have

$$A(\operatorname{adj} A) = (\det(A))I_n = (\operatorname{adj} A)A.$$

*Proof.* For  $1 \leq i \leq n$ , the  $i, i$ -entry of the product  $A(\operatorname{adj} A)$  is

$$\sum_{k=1}^n A(i, k)[\operatorname{adj} A](k, i) = \sum_{k=1}^n (-1)^{i+k} A(i, k) \det(A[i|k]) = \det(A),$$

by Laplace expansion along row  $i$  of  $A$ . Now suppose  $i \neq j$ . The  $i, j$ -entry of  $A(\operatorname{adj} A)$  is

$$\sum_{k=1}^n A(i, k)[\operatorname{adj} A](k, j) = \sum_{k=1}^n (-1)^{j+k} A(i, k) \det(A[j|k]).$$

Let  $C$  be the matrix obtained from  $A$  by replacing row  $j$  of  $A$  by row  $i$  of  $A$ . Then  $C(j, k) = A(i, k)$  and  $C[j|k] = A[j|k]$  for all  $k$ . So the preceding expression is the Laplace expansion for  $\det(C)$  along row  $j$ . On the other hand,  $\det(C) = 0$  because  $C$  has two equal rows. So  $[A(\operatorname{adj} A)](i, j) = 0$ . We have proved that  $A(\operatorname{adj} A)$  is a diagonal matrix with all diagonal entries equal to  $\det(A)$ , as desired. The analogous result for  $(\operatorname{adj} A)A$  is proved similarly, using column expansions.  $\square$

**9.51. Corollary: Formula for the Inverse of a Matrix.** If  $A \in M_n(R)$  and  $\det(A)$  is an invertible element of  $R$ , then the matrix  $A$  is invertible in  $M_n(R)$  with inverse

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj} A.$$

**9.52. Remark.** Conversely, if  $A$  is invertible in  $M_n(R)$ , then  $\det(A)$  is an invertible element of  $R$ . The proof uses the following *product formula for determinants*:

$$\det(AB) = \det(A)\det(B) = \det(B)\det(A) \quad (A, B \in M_n(R)).$$

Taking  $B = A^{-1}$ , the left side becomes  $\det(I_n) = 1_R$ , so  $\det(B)$  is a two-sided inverse of  $\det(A)$  in  $R$ . We will deduce the product formula as a consequence of the Cauchy-Binet formula, which is proved in the next section.

## 9.7 Cauchy-Binet Formula

This section discusses the Cauchy-Binet formula, which expresses the determinant of a product of rectangular matrices as the sum of a product of determinants of suitable submatrices. The proof of this formula is a nice application of the properties of inversions and determinants.

To state the Cauchy-Binet formula, we need the following notation. Given a  $c \times d$  matrix  $M$ , write  $M_i$  for the  $i$ th row of  $M$  and  $M^j$  for the  $j$ th column of  $M$ . Given indices  $j_1, \dots, j_c \in \{1, 2, \dots, d\}$ , let  $(M^{j_1}, \dots, M^{j_c})$  denote the  $c \times c$  matrix whose columns are  $M^{j_1}, \dots, M^{j_c}$  in this order. Similarly,  $(M_{i_1}, \dots, M_{i_d})$  is the matrix whose rows are  $M_{i_1}, \dots, M_{i_d}$  in this order.

**9.53. Theorem: Cauchy-Binet Formula.** Suppose  $m \leq n$ ,  $A$  is an  $m \times n$  matrix, and  $B$  is an  $n \times m$  matrix. Let  $J$  be the set of all lists  $j = (j_1, j_2, \dots, j_m)$  such that  $1 \leq j_1 < j_2 < \dots < j_m \leq n$ . Then

$$\det(AB) = \sum_{j \in J} \det(A^{j_1}, A^{j_2}, \dots, A^{j_m}) \det(B_{j_1}, B_{j_2}, \dots, B_{j_m}).$$

*Proof.* Note that all matrices appearing in the formula are  $m \times m$ , so all the determinants are defined. We begin by using the definitions of matrix products and determinants (§9.5) to write

$$\det(AB) = \sum_{w \in S_m} \operatorname{sgn}(w) \prod_{i=1}^m (AB)(i, w(i)) = \sum_{w \in S_m} \operatorname{sgn}(w) \prod_{i=1}^m \sum_{k_i=1}^n A(i, k_i) B(k_i, w(i)).$$

The generalized distributive law (§2.1) changes the product of sums into a sum of products:

$$\det(AB) = \sum_{w \in S_m} \sum_{k_1=1}^n \cdots \sum_{k_m=1}^n \operatorname{sgn}(w) \prod_{i=1}^m A(i, k_i) \prod_{i=1}^m B(k_i, w(i)).$$

Let  $K$  be the set of all lists  $k = (k_1, \dots, k_m)$  with every  $k_i \in \{1, 2, \dots, n\}$ , and let  $K'$  be the set of lists in  $K$  whose entries  $k_i$  are distinct. We can combine the  $m$  separate sums over the  $k_i$ 's into a single sum over lists  $k \in K$ . We can also reorder the summation to get

$$\det(AB) = \sum_{k \in K} \sum_{w \in S_m} \operatorname{sgn}(w) \prod_{i=1}^m A(i, k_i) \prod_{i=1}^m B(k_i, w(i)).$$

Next, factor out quantities that do not depend on  $w$ :

$$\det(AB) = \sum_{k \in K} \prod_{i=1}^m A(i, k_i) \left[ \sum_{w \in S_m} \operatorname{sgn}(w) \prod_{i=1}^m B(k_i, w(i)) \right].$$

The term in brackets is the defining formula for  $\det(B_{k_1}, \dots, B_{k_m})$ . If any two entries in  $(k_1, \dots, k_m)$  are equal, this matrix has two equal rows, so its determinant is zero. Discarding these terms, we are reduced to summing over lists  $k \in K'$ . So now we have

$$\det(AB) = \sum_{k \in K'} \prod_{i=1}^m A(i, k_i) \det(B_{k_1}, \dots, B_{k_m}).$$

To continue, observe that for every list  $k \in K'$  there exists a unique sorted list  $j \in J$  with  $j = \operatorname{sort}(k)$ . Grouping summands gives

$$\det(AB) = \sum_{j \in J} \sum_{\substack{k \in K' \\ \operatorname{sort}(k)=j}} \prod_{i=1}^m A(i, k_i) \det(B_{k_1}, \dots, B_{k_m}).$$

Given that  $\operatorname{sort}(k) = j$ , we can change the matrix  $(B_{k_1}, \dots, B_{k_m})$  into the matrix  $(B_{j_1}, \dots, B_{j_m})$  by repeatedly switching adjacent rows. Each such switch flips the sign of the determinant, and the number of row switches required is readily seen to be  $\operatorname{inv}(k_1 k_2 \cdots k_m)$ . (To see this, adapt the proof of 9.29 to the case where the objects being sorted are  $\{j_1 < j_2 < \dots < j_m\}$  instead of  $\{1 < 2 < \dots < m\}$ ; cf. 9.179.) Letting  $\operatorname{sgn}(k) = (-1)^{\operatorname{inv}(k)}$ , we can therefore write

$$\det(AB) = \sum_{j \in J} \sum_{\substack{k \in K' \\ \operatorname{sort}(k)=j}} \operatorname{sgn}(k) \prod_{i=1}^m A(i, k_i) \det(B_{j_1}, \dots, B_{j_m}).$$

The determinant in this formula depends only on  $j$ , not on  $k$ , so it can be brought out of the inner summation:

$$\det(AB) = \sum_{j \in J} \det(B_{j_1}, \dots, B_{j_m}) \sum_{\substack{k \in K' \\ \text{sort}(k)=j}} \text{sgn}(k) \prod_{i=1}^m A(i, k_i).$$

To finish, note that every  $k \in K'$  that sorts to  $j$  can be written as  $(k_1, \dots, k_m) = (j_{v(1)}, \dots, j_{v(m)})$  for a uniquely determined permutation  $v \in S_m$ . Since  $j$  is an increasing sequence, it follows that  $\text{inv}(k) = \text{inv}(v)$  and  $\text{sgn}(k) = \text{sgn}(v)$ . Changing variables in the inner summation, we get

$$\det(AB) = \sum_{j \in J} \det(B_{j_1}, \dots, B_{j_m}) \left[ \sum_{v \in S_m} \text{sgn}(v) \prod_{i=1}^m A(i, j_{v(i)}) \right].$$

The term in brackets is none other than  $\det(A^{j_1}, \dots, A^{j_m})$ , so the proof is complete.  $\square$

**9.54. Theorem: Product Formula for Determinants.** If  $A$  and  $B$  are  $m \times m$  matrices, then  $\det(AB) = \det(A) \det(B)$ .

*Proof.* Take  $n = m$  in the Cauchy-Binet formula. The index set  $J$  consists of the single list  $(1, 2, \dots, m)$ , and the summand corresponding to this list reduces to  $\det(A) \det(B)$ .  $\square$

Other examples of combinatorial proofs of determinant formulas appear in §11.14 and §12.9.

## 9.8 Subgroups

Suppose  $(G, \star)$  is a group, and  $H$  is a subset of  $G$ . One might hope that  $(H, \star')$  is also a group, where  $\star'$  is the restriction of  $\star$  to  $H \times H$ . This is not true in general, but it will be true if  $H$  is a *subgroup*.

**9.55. Definition: Subgroups.** Let  $(G, \star)$  be a group and let  $H$  be a subset of  $G$ .  $H$  is called a *subgroup* of  $G$ , written  $H \leq G$ , iff the following three “closure conditions” are satisfied:

$$\begin{array}{ll} e_G \in H & \text{(closure under identity);} \\ \forall a, b \in H, a \star b \in H & \text{(closure under the operation);} \\ \forall a \in H, a^{-1} \in H & \text{(closure under inverses).} \end{array}$$

A subgroup  $H$  is called *normal* in  $G$ , written  $H \trianglelefteq G$ , iff

$$\forall g \in G, \forall h \in H, ghg^{-1} \in H \quad \text{(closure under conjugation).}$$

Let us verify that  $(H, \star')$  is indeed a group when  $H \leq G$ . Since  $H$  is closed under the operation,  $\star'$  does map  $H \times H$  into  $H$  (not just into  $G$ ), so the closure axiom holds for  $(H, \star')$ . Since  $H$  is a subset of  $G$ , associativity holds in  $H$  because it is known to hold in  $G$ . The identity  $e$  of  $G$  lies in  $H$  by assumption. Since  $e \star h = h = h \star e$  holds for all  $h \in G$ , the relation  $e \star' h = h = h \star' e$  certainly holds for all  $h \in H \subseteq G$ . Finally, every element  $x$  of  $H$  has an inverse  $y$  (relative to  $\star$ ) that lies in  $H$ , by assumption. Now  $y$  is still an inverse of  $x$  relative to  $\star'$ , so the proof is complete. One also sees that  $H$  is commutative if  $G$  is commutative, but the converse statement is not always true. Usually, we use the same symbol  $\star$  (instead of  $\star'$ ) to denote the operation in the subgroup  $H$ .

**9.56. Example.** We have the following chain of subgroups of the additive group  $(\mathbb{C}, +)$ :

$$\{0\} \leq \mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}.$$

Similarly,  $\{-1, 1\}$  and  $\mathbb{Q}^+$  are both subgroups of  $(\mathbb{Q} \sim \{0\}, \times)$ . The set  $\{0, 3, 6, 9\}$  is a subgroup of  $(\mathbb{Z}_{12}, \oplus)$ ; one can prove closure under addition and inverses by a finite case analysis, or by inspection of the relevant portion of the addition table for  $\mathbb{Z}_{12}$ .

**9.57. Example.** The sets  $H = \{(1)(2)(3), (1, 2, 3), (1, 3, 2)\}$  and  $K = \{(1)(2)(3), (1, 3)\}$  are subgroups of  $S_3$ , as one readily verifies. Moreover,  $H$  is normal in  $S_3$ , but  $K$  is not. The set  $J = \{(1)(2)(3), (1, 3), (2, 3), (1, 3)\}$  is not a subgroup of  $S_3$ , since closure under the operation fails:  $(1, 3) \circ (2, 3) = (1, 3, 2) \notin J$ . Here is a four-element normal subgroup of  $S_4$ :

$$V = \{(1)(2)(3)(4), (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\}.$$

Each element of  $V$  is its own inverse, and one confirms closure of  $V$  under the operation by checking all possible products. To prove the normality of  $V$  in  $S_4$ , it is helpful to use 9.131 below.

**9.58. Example.** The set of *even integers* is a subgroup of  $(\mathbb{Z}, +)$ . For, the identity element zero is even; the sum of two even integers is again even; and  $x$  even implies  $-x$  is even. More generally, let  $k$  be any fixed integer, and let  $H = \{kn : n \in \mathbb{Z}\}$  consist of all integer multiples of  $k$ . A routine verification shows that  $H$  is a subgroup of  $(\mathbb{Z}, +)$ . We write  $H = k\mathbb{Z}$  for brevity. The next theorem shows that we have found *all* the subgroups of the additive group  $\mathbb{Z}$ .

**9.59. Theorem: Subgroups of  $\mathbb{Z}$ .** Every subgroup  $H$  of  $(\mathbb{Z}, +)$  has the form  $k\mathbb{Z}$  for a unique integer  $k \geq 0$ .

*Proof.* We have noted that all the subsets  $k\mathbb{Z}$  are indeed subgroups. Given an arbitrary subgroup  $H$ , consider two cases. If  $H = \{0\}$ , then  $H = 0\mathbb{Z}$ . Otherwise,  $H$  contains at least one nonzero integer  $m$ . If  $m$  is negative, then  $-m \in H$  since  $H$  is closed under inverses. So,  $H$  contains strictly positive integers. Take  $k$  to be the least positive integer in  $H$ . We claim that  $H = k\mathbb{Z}$ . Let us prove that  $kn \in H$  for all  $n \in \mathbb{Z}$ , so that  $k\mathbb{Z} \subseteq H$ . For  $n \geq 0$ , we argue by induction on  $n$ . When  $n = 0$ , we must prove  $k0 = 0 \in H$ , which holds since  $H$  contains the identity of  $\mathbb{Z}$ . When  $n = 1$ , we must prove  $k1 = k \in H$ , which is true by choice of  $k$ . Assume  $n \geq 1$  and  $kn \in H$ . Then  $k(n+1) = kn + k \in H$  since  $kn \in H$ ,  $k \in H$ , and  $H$  is closed under addition. Finally, for negative  $n$ , write  $n = -m$  and note that  $kn = -(km) \in H$  since  $km \in H$  and  $H$  is closed under inverses.

The key step is to prove the reverse inclusion  $H \subseteq k\mathbb{Z}$ . Fix  $z \in H$ . Dividing  $z$  by  $k$ , we obtain  $z = kq + r$  for some integers  $q, r$  with  $0 \leq r < k$ . By what we proved in the last paragraph,  $k(-q) \in H$ . So,  $r = z - kq = z + k(-q) \in H$  since  $H$  is closed under addition. Now, since  $k$  is the *least* positive integer in  $H$ , we cannot have  $0 < r < k$ . The only possibility left is  $r = 0$ , so  $z = kq \in k\mathbb{Z}$ , as desired.

Finally, to prove uniqueness, suppose  $k\mathbb{Z} = m\mathbb{Z}$  for  $k, m \geq 0$ . Note  $k = 0$  iff  $m = 0$ , so assume  $k, m > 0$ . Since  $k \in k\mathbb{Z} = m\mathbb{Z}$ ,  $k$  is a multiple of  $m$ . Similarly,  $m$  is a multiple of  $k$ . As both  $k$  and  $m$  are positive, this forces  $k = m$ , completing the proof.  $\square$

How can we find subgroups of a given group  $G$ ? As we see next, each element  $x \in G$  gives rise to a subgroup of  $G$  in a natural way.

**9.60. Definition: Cyclic Subgroups and Cyclic Groups.** Let  $G$  be a group written multiplicatively, and let  $x \in G$ . The *cyclic subgroup of  $G$  generated by  $x$*  is  $\langle x \rangle = \{x^n : n \in \mathbb{Z}\}$ . One sees, using the laws of exponents, that this subset of  $G$  really is a subgroup.  $G$  is called a *cyclic group* iff there exists  $x \in G$  with  $G = \langle x \rangle$ . When  $G$  is written additively, we have  $\langle x \rangle = \{nx : n \in \mathbb{Z}\}$ .



**9.61. Example.** The group  $(\mathbb{Z}, +)$  is cyclic, since  $\mathbb{Z} = \langle 1 \rangle = \langle -1 \rangle$ . The subgroups  $k\mathbb{Z} = \langle k \rangle$  considered above are cyclic subgroups of  $\mathbb{Z}$ . Our last theorem implies that *every subgroup of  $\mathbb{Z}$  is cyclic*. The groups  $(\mathbb{Z}_n, \oplus)$  are also cyclic; each of these groups is generated by 1. The group  $(\{a, b, c, d\}, \star)$  discussed at the end of 9.7 is *not* cyclic. To prove this, we compute all the cyclic subgroups of this group:

$$\langle a \rangle = \{a\}, \quad \langle b \rangle = \{a, b\}, \quad \langle c \rangle = \{a, c\}, \quad \langle d \rangle = \{a, d\}.$$

None of the cyclic subgroups equals the whole group, so the group is not cyclic. For a bigger example of a non-cyclic group, consider  $(\mathbb{Q}, +)$ . Any nonzero cyclic subgroup has the form  $\langle a/b \rangle$  for some positive rational number  $a/b$ . One may check that  $a/2b$  does not lie in this subgroup, so  $\mathbb{Q} \neq \langle a/b \rangle$ . Noncommutative groups furnish additional examples of non-cyclic groups, as the next result shows.

**9.62. Theorem: Cyclic Groups are Commutative.**

*Proof.* Let  $G = \langle x \rangle$  be cyclic. Given  $y, z \in G$ , we can write  $y = x^n$  and  $z = x^m$  for some  $n, m \in \mathbb{Z}$ . Since integer addition is commutative, the laws of exponents give

$$yz = x^n x^m = x^{n+m} = x^{m+n} = x^m x^n = zy. \quad \square$$

By adapting the argument in 9.59, one can show that *every subgroup of a cyclic group is cyclic*; we leave this as an exercise for the reader.

**9.63. Example.** The cyclic group  $\mathbb{Z}_6$  has the following cyclic subgroups (which are *all* the subgroups of this group):

$$\langle 0 \rangle = \{0\}; \quad \langle 1 \rangle = \{0, 1, 2, 3, 4, 5\} = \langle 5 \rangle; \quad \langle 2 \rangle = \{0, 2, 4\} = \langle 4 \rangle; \quad \langle 3 \rangle = \{0, 3\}.$$

In the group  $S_4$ , we have

$$\langle (1, 3, 4, 2) \rangle = \{(1, 3, 4, 2), (1, 4)(3, 2), (1, 2, 4, 3), (1)(2)(3)(4)\}.$$

## 9.9 Automorphism Groups of Graphs

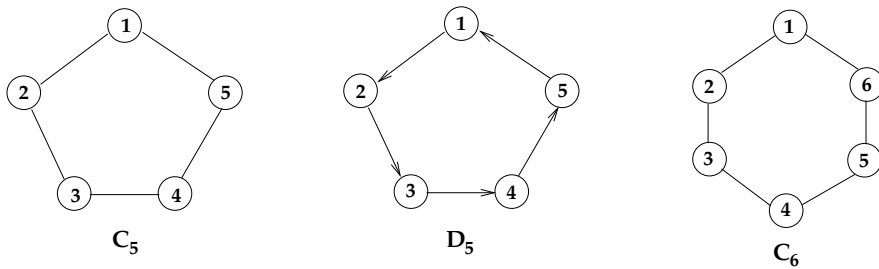
This section uses graphs to construct examples of subgroups of symmetric groups. These subgroups will be used later when we discuss applications of group theory to counting problems.

**9.64. Definition: Automorphism Group of a Graph.** Let  $K$  be a simple graph with vertex set  $X$  and edge set  $E$ . A *graph automorphism* of  $K$  is a bijection  $f : X \rightarrow X$  such that, for all  $u \neq v$  in  $X$ ,  $\{u, v\} \in E$  iff  $\{f(u), f(v)\} \in E$ . Let  $\text{Aut}(K)$  denote the set of all graph automorphisms of  $K$ . Analogous definitions are made for directed simple graphs; here, the requirement on  $f$  is that  $(u, v) \in E$  iff  $(f(u), f(v)) \in E$  for all  $u, v \in X$ .

One verifies immediately from the definition that  $\text{Aut}(K) \leq (\text{Sym}(X), \circ)$ . Thus, automorphism groups of graphs are subgroups of symmetric groups.

**9.65. Example.** Consider the graphs shown in Figure 9.2. The undirected cycle  $C_5$  has exactly ten automorphisms. They are given in one-line form in the following list:

$$\begin{array}{cccccc} [1\ 2\ 3\ 4\ 5], & [2\ 3\ 4\ 5\ 1], & [3\ 4\ 5\ 1\ 2], & [4\ 5\ 1\ 2\ 3], & [5\ 1\ 2\ 3\ 4], \\ [5\ 4\ 3\ 2\ 1], & [4\ 3\ 2\ 1\ 5], & [3\ 2\ 1\ 5\ 4], & [2\ 1\ 5\ 4\ 3], & [1\ 5\ 4\ 3\ 2]. \end{array}$$


**FIGURE 9.2**

Graphs used to illustrate automorphism groups.

The same automorphisms, written in cycle notation, look like this:

$$(1)(2)(3)(4)(5), \quad (1, 2, 3, 4, 5), \quad (1, 3, 5, 2, 4), \quad (1, 4, 2, 5, 3), \quad (1, 5, 4, 3, 2), \\ (1, 5)(2, 4)(3), \quad (1, 4)(2, 3)(5), \quad (1, 3)(4, 5)(2), \quad (1, 2)(3, 5)(4), \quad (2, 5)(3, 4)(1).$$

Geometrically, we can think of  $C_5$  as a *necklace* with five beads. The first five automorphisms on each list arise by rotating the necklace through various angles (rotation by zero is the identity map). The next five automorphisms arise by reflecting the necklace in five possible axes of symmetry.

Now consider the automorphism group of the *directed* cycle  $D_5$ . Every automorphism of the directed graph  $D_5$  is automatically an automorphism of the associated undirected graph  $C_5$ , so  $\text{Aut}(D_5) \leq \text{Aut}(C_5)$ . However, not every automorphism of  $C_5$  is an automorphism of  $D_5$ . In this example, the five “rotations” preserve the direction of the edges, hence are automorphisms of  $D_5$ . But the five “reflections” reverse the direction of the edges, so these are not elements of  $\text{Aut}(D_5)$ . We can write  $\text{Aut}(D_5) = \langle (1, 2, 3, 4, 5) \rangle$ , so that this automorphism group is cyclic of size 5.

The 6-cycle  $C_6$  can be analyzed in a similar way. The automorphism group consists of six “rotations” and six “reflections,” which are given in cycle form below:

$$(1)(2)(3)(4)(5)(6), \quad (1, 2, 3, 4, 5, 6), \quad (1, 3, 5)(2, 4, 6), \quad (1, 4)(2, 5)(3, 6), \\ (1, 5, 3)(2, 6, 4), \quad (1, 6, 5, 4, 3, 2), \quad (2, 6)(3, 5)(1, 4), \quad (1, 2)(3, 6)(4, 5), \\ (1, 3)(4, 6)(2, 5), \quad (2, 3)(5, 6)(1, 4), \quad (1, 5)(2, 4)(3, 6), \quad (1, 6)(2, 5)(3, 4).$$

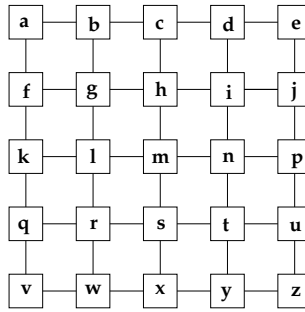
The observations in the previous example generalize as follows.

**9.66. Theorem: Automorphism Group of a Cycle.** For  $n \geq 3$ , let  $C_n$  be the graph with vertex set  $X = \{1, 2, \dots, n\}$  and edge set  $E = \{\{i, i+1\} : 1 \leq i < n\} \cup \{\{1, n\}\}$ . Then  $\text{Aut}(C_n)$  is a subgroup of  $S_n$  of size  $2n$ . The elements of this group (in one-line form) are the  $n$  permutations

$$[i, i+1, i+2, \dots, n, 1, 2, \dots, i-1] \quad (1 \leq i \leq n) \quad (9.5)$$

together with the reversals of these  $n$  words.

*Proof.* It is routine to check that all of the displayed permutations do preserve the edges of  $C_n$ , hence are automorphisms of this graph. We must show that these are the *only* automorphisms of  $C_n$ . Let  $g$  be any automorphism of  $C_n$ , and put  $i = g(1)$ . Now, since 1 and 2 are adjacent in  $C_n$ ,  $g(1)$  and  $g(2)$  must also be adjacent in  $C_n$ . There are two cases:  $g(2) = i+1$  or  $g(2) = i-1$  (reading values mod  $n$ ). Suppose the first case occurs. Since 2 is adjacent to 3, we must have  $g(3) = i+2$  or  $g(3) = i$ . But  $i = g(1)$  and  $g$  is injective, so it

**FIGURE 9.3**

Graph representing a  $5 \times 5$  chessboard.

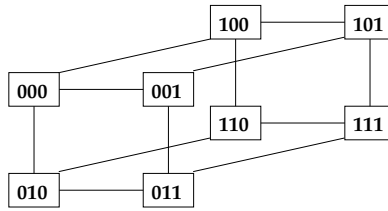
must be that  $g(3) = i + 2$ . Continuing around the cycle in this way, we see that  $g$  must be one of the permutations displayed in (9.5). Similarly, in the case where  $g(2) = i - 1$ , we see that  $g(3) = i - 2$ , etc., and  $g$  must be the reversal of one of the permutations in (9.5).  $\square$

The reasoning used in the preceding proof can be adapted to determine the automorphism groups of more complicated graphs.

**9.67. Example.** Consider the graph  $B$  displayed in Figure 9.3, which models a  $5 \times 5$  chessboard. What are the automorphisms of  $B$ ? We note that  $B$  has four vertices of degree 2:  $a$ ,  $e$ ,  $v$ , and  $z$ . An automorphism  $\phi$  of  $B$  must restrict to give a permutation of these four vertices, since automorphisms preserve degree. Suppose, for example, that  $\phi(a) = v$ . What can  $\phi(b)$  be in this situation? Evidently,  $\phi(b)$  must be  $q$  or  $w$ . In the former case,  $\phi(c) = k$  is forced by degree considerations, whereas  $\phi(c) = x$  is forced in the latter case. Continuing around the “edge” of the graph, we see that the action of  $\phi$  on all of the “border” vertices is completely determined by where  $a$  and  $b$  go. A tedious but routine argument then shows that the images of the remaining vertices are also forced. Since  $a$  can map to one of the four corners, and then  $b$  can map to one of the two neighbors of  $\phi(a)$ , there are at most  $4 \times 2 = 8$  automorphisms of  $B$ . Here are the eight possibilities in cycle form:

$$\begin{aligned}
 r_0 &= (a)(b)(c) \cdots (x)(y)(z) = \text{id}; \\
 r_1 &= (a, e, z, v)(b, j, y, q)(c, p, x, k)(d, u, w, f)(g, i, t, r)(h, n, s, l)(m); \\
 r_2 &= (a, z)(b, y)(c, x)(d, w)(e, v)(j, q)(p, k)(u, f)(g, t)(h, s)(i, r)(n, l)(m); \\
 r_3 &= (a, v, z, e)(b, q, y, j)(c, k, x, p)(d, f, w, u)(g, r, t, i)(h, l, s, n)(m); \\
 s_{y=0} &= (a, v)(b, w)(c, x)(d, y)(e, z)(f, q)(g, r)(h, s)(i, t)(j, u)(k, l)(m)(n)(p); \\
 s_{x=0} &= (a, e)(b, d)(f, j)(g, i)(k, p)(l, n)(q, u)(r, t)(v, z)(w, y)(c, h)(m)(s)(x); \\
 s_{y=x} &= (a, z)(b, u)(c, p)(d, j)(f, y)(g, t)(h, n)(k, x)(l, s)(q, w)(e, i)(m)(r)(v); \\
 s_{y=-x} &= (b, f)(c, k)(d, q)(e, v)(h, l)(i, r)(j, w)(n, s)(p, x)(u, y)(a, g)(m)(t)(z).
 \end{aligned}$$

One may check that all of these maps really are automorphisms of  $B$ , so  $|\text{Aut}(B)| = 8$ . The reader will perceive that this graph has essentially the same symmetries as  $C_4$ : four rotations and four reflections. (The subscripts of the reflections indicate the axis of reflection, taking  $m$  to be located at the origin.) By directing the edges in a suitable way, we could produce a graph with only four automorphisms (the rotations). These graphs and groups will play a crucial role in solving the chessboard-coloring problem mentioned in the introduction to this chapter.


**FIGURE 9.4**

The cube graph.

**9.68. Example.** As a final illustration of the calculation of an automorphism group, consider the graph  $C$  shown in Figure 9.4, which models a three-dimensional cube. We have taken the vertex set of  $C$  to be  $\{0, 1\}^3$ , the set of binary words of length 3. Which bijections  $f : \{0, 1\}^3 \rightarrow \{0, 1\}^3$  might be automorphisms of  $C$ ? First,  $f(000)$  can be any of the eight vertices. Next, the three neighbors of 000 (namely 001, 010, and 100) can be mapped bijectively onto the three neighbors of  $f(000)$  in any of  $3! = 6$  ways. The images of the remaining four vertices are now uniquely determined, as one may check. By the product rule, there are at most  $8 \times 6 = 48$  automorphisms of  $C$ . A routine but tedious verification shows that all of these potential automorphisms really are automorphisms, so  $|\text{Aut}(C)| = 48$ . The geometrically inclined reader may like to visualize these automorphisms as arising from suitable rotations and reflections in three-dimensional space. Here are the six automorphisms of  $C$  that send 000 to 110:

$$\begin{aligned}
 f_1 &= \begin{pmatrix} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ 110 & 100 & 111 & 101 & 010 & 000 & 011 & 001 \end{pmatrix}, \\
 f_2 &= \begin{pmatrix} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ 110 & 100 & 010 & 000 & 111 & 101 & 011 & 001 \end{pmatrix}, \\
 f_3 &= \begin{pmatrix} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ 110 & 111 & 100 & 101 & 010 & 011 & 000 & 001 \end{pmatrix}, \\
 f_4 &= \begin{pmatrix} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ 110 & 111 & 010 & 011 & 100 & 101 & 000 & 001 \end{pmatrix}, \\
 f_5 &= \begin{pmatrix} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ 110 & 010 & 100 & 000 & 111 & 011 & 101 & 001 \end{pmatrix}, \\
 f_6 &= \begin{pmatrix} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ 110 & 010 & 111 & 011 & 100 & 000 & 101 & 001 \end{pmatrix}.
 \end{aligned}$$

## 9.10 Group Homomorphisms

**9.69. Definition: Group Homomorphisms.** Let  $(G, \star)$  and  $(H, \bullet)$  be groups. A function  $f : G \rightarrow H$  is called a *group homomorphism* iff

$$f(x \star y) = f(x) \bullet f(y) \quad \text{for all } x, y \in G.$$

A *group isomorphism* is a bijective group homomorphism.

**9.70. Example.** Define  $f : \mathbb{R} \rightarrow \mathbb{R}^+$  by  $f(x) = e^x$ . This function is a group homomorphism from  $(\mathbb{R}, +)$  to  $(\mathbb{R}^+, \times)$ , since  $f(x+y) = e^{x+y} = e^x \times e^y = f(x) \times f(y)$  for all  $x, y \in \mathbb{R}$ . In fact,  $f$  is a group isomorphism since  $g : \mathbb{R}^+ \rightarrow \mathbb{R}$  given by  $g(x) = \ln x$  is a two-sided inverse for  $f$ .

**9.71. Example.** Define  $h : \mathbb{C} \rightarrow \mathbb{R}$  by  $h(x+iy) = x$  for all  $x, y \in \mathbb{R}$ . One checks that  $h$  is a group homomorphism from  $(\mathbb{C}, +)$  to  $(\mathbb{R}, +)$  that is surjective but not injective. Next, define  $r : \mathbb{C} \setminus \{0\} \rightarrow \mathbb{R} \setminus \{0\}$  by setting  $r(x+iy) = |x+iy| = \sqrt{x^2+y^2}$ . Given nonzero  $w = x+iy$  and  $z = u+iv$ , we calculate

$$\begin{aligned} r(wz) &= r((xu-yv) + i(yu+xv)) = \sqrt{(xu-yv)^2 + (yu+xv)^2} \\ &= \sqrt{(x^2+y^2)(u^2+v^2)} = r(w)r(z). \end{aligned}$$

So  $r$  is a homomorphism of multiplicative groups.

**9.72. Example.** For any group  $G$ , the identity map  $\text{id}_G : G \rightarrow G$  is a group isomorphism. More generally, if  $H \leq G$ , then the inclusion map  $j : H \rightarrow G$  given by  $j(h) = h$  for  $h \in H$  is a group homomorphism. If  $f : G \rightarrow K$  and  $g : K \rightarrow P$  are group homomorphisms, then  $g \circ f : G \rightarrow P$  is a group homomorphism, since

$$(g \circ f)(xy) = g(f(xy)) = g(f(x)f(y)) = g(f(x))g(f(y)) = (g \circ f)(x)(g \circ f)(y) \quad (x, y \in G).$$

Moreover,  $g \circ f$  is an isomorphism if  $f$  and  $g$  are isomorphisms, since the composition of bijections is a bijection. If  $f : G \rightarrow K$  is an isomorphism, then  $f^{-1} : K \rightarrow G$  is also an isomorphism. For suppose  $u, v \in K$ . Write  $x = f^{-1}(u)$  and  $y = f^{-1}(v)$ , so  $u = f(x)$  and  $v = f(y)$ . Since  $f$  is a group homomorphism, it follows that  $uv = f(xy)$ . Applying  $f^{-1}$  to this relation, we get  $f^{-1}(uv) = xy = f^{-1}(u)f^{-1}(v)$ .

**9.73. Definition: Automorphism Groups.** Let  $(G, \star)$  be a group. An *automorphism* of  $G$  is a group isomorphism  $f : G \rightarrow G$ . Let  $\text{Aut}(G)$  denote the set of all such automorphisms.

The remarks in the preceding example (with  $K = P = G$ ) show that  $\text{Aut}(G)$  is a subgroup of  $(\text{Sym}(G), \circ)$ .

**9.74. Example: Inner Automorphisms.** Let  $G$  be any group, and fix an element  $g \in G$ . Define a map  $C_g : G \rightarrow G$  (called *conjugation by  $g$* ) by setting  $C_g(x) = gxg^{-1}$ . This map is a group homomorphism, since  $C_g(xy) = g(xy)g^{-1} = (gxg^{-1})(gyg^{-1}) = C_g(x)C_g(y)$  for all  $x, y \in G$ . Furthermore,  $C_g$  is a group isomorphism, since a calculation shows that  $C_{g^{-1}}$  is the two-sided inverse of  $C_g$ . It follows that  $C_g \in \text{Aut}(G)$  for every  $g \in G$ . We call automorphisms of the form  $C_g$  *inner automorphisms of  $G$* . It is possible for different group elements to induce the same inner automorphism of  $G$ . For example, if  $G$  is commutative, then  $C_g(x) = gxg^{-1} = gg^{-1}x = x$  for all  $g, x \in G$ , so that all of the conjugation maps  $C_g$  reduce to  $\text{id}_G$ .

**9.75. Theorem: Properties of Group Homomorphisms.** Let  $f : G \rightarrow H$  be a group homomorphism. For all  $n \in \mathbb{Z}$  and all  $x \in G$ ,  $f(x^n) = f(x)^n$ . In particular,  $f(e_G) = e_H$  and  $f(x^{-1}) = f(x)^{-1}$ . We say that  $f$  *preserves powers, identities, and inverses*.

*Proof.* First we prove the result for all  $n \geq 0$  by induction on  $n$ . When  $n = 0$ , we must prove that  $f(e_G) = e_H$ . Note that  $e_G e_G = e_G$ . Applying  $f$  to both sides of this equation gives

$$f(e_G)f(e_G) = f(e_G e_G) = f(e_G) = f(e_G)e_H.$$

By left cancellation of  $f(e_G)$  in  $H$ , we conclude that  $f(e_G) = e_H$ . For the induction step,

assume  $n \geq 0$  and  $f(x^n) = f(x)^n$ ; we will prove  $f(x^{n+1}) = f(x)^{n+1}$ . Using the definition of exponent notation, we calculate

$$f(x^{n+1}) = f(x^n x) = f(x^n) f(x) = f(x)^n f(x) = f(x)^{n+1}.$$

Next, let us prove the result when  $n = -1$ . Given  $x \in G$ , apply  $f$  to the equation  $xx^{-1} = e_G$  to obtain

$$f(x)f(x^{-1}) = f(xx^{-1}) = f(e_G) = e_H = f(x)f(x)^{-1}.$$

Left cancellation of  $f(x)$  gives  $f(x^{-1}) = f(x)^{-1}$ . Finally, consider an arbitrary negative integer  $n = -m$ , where  $m > 0$ . We have

$$f(x^n) = f((x^m)^{-1}) = f(x^m)^{-1} = (f(x)^m)^{-1} = f(x)^{-m} = f(x)^n. \quad \square$$

We can use group homomorphisms to construct more examples of subgroups.

**9.76. Definition: Kernel and Image of a Homomorphism.** Let  $f : G \rightarrow H$  be a group homomorphism. The *kernel* of  $f$ , denoted  $\ker(f)$ , is the set of all  $x \in G$  such that  $f(x) = e_H$ . The *image* of  $f$ , denoted  $\text{img}(f)$ , is the set of all  $y \in H$  such that  $y = f(z)$  for some  $z \in G$ .

The reader may check that  $\ker(f) \trianglelefteq G$  and  $\text{img}(f) \leq H$ .

**9.77. Example.** Consider the homomorphisms  $h$  and  $r$  from 9.71, given by  $h(x + iy) = x$  and  $r(z) = |z|$  for  $x, y \in \mathbb{R}$  and nonzero  $z \in \mathbb{C}$ . The kernel of  $h$  is the set of pure imaginary numbers  $\{iy : y \in \mathbb{R}\}$ . The kernel of  $r$  is the unit circle  $\{z \in \mathbb{C} : |z| = 1\}$ . The image of  $h$  is all of  $\mathbb{R}$ , while the image of  $r$  is  $\mathbb{R}^+$ .

**9.78. Example: Even Permutations.** By 9.31, the function  $\text{sgn} : S_n \rightarrow \{+1, -1\}$  is a group homomorphism. The kernel of this homomorphism, which is denoted  $A_n$ , consists of all  $f \in S_n$  such that  $\text{sgn}(f) = +1$ . Such  $f$  are called *even permutations*.  $A_n$  is called the *alternating group on  $n$  letters*. We will see later (9.121) that  $|A_n| = |S_n|/2 = n!/2$  for all  $n \geq 2$ .

**9.79. Example: Analysis of Cyclic Subgroups.** Let  $G$  be any group, written multiplicatively, and fix an element  $x \in G$ . Define  $f : \mathbb{Z} \rightarrow G$  by setting  $f(n) = x^n$  for all  $n \in \mathbb{Z}$ . By the laws of exponents,  $f$  is a group homomorphism. The image of  $f$  is precisely  $\langle x \rangle$ , the cyclic subgroup of  $G$  generated by  $x$ . The kernel of  $f$  is some subgroup of  $\mathbb{Z}$ , which by 9.59 has the form  $m\mathbb{Z}$  for some integer  $m \geq 0$ . Consider the case where  $m = 0$ . Then  $x^i = e_G$  iff  $f(i) = e_G$  iff  $i \in \ker(f)$  iff  $i = 0$ , so  $x^0$  is the only power of  $x$  that equals the identity of  $G$ . We say that  $x$  has *infinite order* in this case. We remark that  $i \neq j$  implies  $x^i \neq x^j$ , since

$$x^i = x^j \Rightarrow x^{i-j} = e_G \Rightarrow i - j = 0 \Rightarrow i = j.$$

In other words, all integer powers of  $x$  are distinct elements of  $G$ . This means that  $f : \mathbb{Z} \rightarrow G$  is injective. So  $f$  induces a group isomorphism  $f' : \mathbb{Z} \rightarrow \langle x \rangle$ .

Now consider the case where  $m > 0$ . Then  $x^i = e_G$  iff  $f(i) = e_G$  iff  $i \in \ker(f)$  iff  $i$  is a multiple of  $m$ . We say that  $x$  has *order  $m$*  in this case; thus, the order of  $x$  is the least positive exponent  $i$  such that  $x^i = e_G$ . We claim that the cyclic group  $\langle x \rangle$  consists of the  $m$  *distinct* elements  $x^0, x^1, x^2, \dots, x^{m-1}$ . For, given an arbitrary element  $x^n \in \langle x \rangle$ , we can divide  $n$  by  $m$  to get  $n = mq + r$  for some  $r$  with  $0 \leq r < m$ . Then  $x^n = x^{mq+r} = (x^m)^q x^r = e_G^q x^r = x^r$ , so  $x^n$  is equal to one of the elements in our list. Furthermore, the listed elements are distinct. For suppose  $0 \leq i < j < m$  and  $x^i = x^j$ . Then  $x^{j-i} = e_G$ , forcing  $m$  to divide  $j - i$ . But  $0 \leq j - i < m$ , so the only possibility is  $j - i = 0$ , hence  $i = j$ . Consider the function  $g : \mathbb{Z}_m \rightarrow \langle x \rangle$  given by  $g(i) = x^i$  for  $0 \leq i < m$ . This function is a well-defined bijection

by the preceding remarks. Furthermore,  $g$  is a group homomorphism. To check this, let  $i, j \in \mathbb{Z}_m$ . If  $i + j < m$ , then

$$g(i \oplus j) = g(i + j) = x^{i+j} = x^i x^j = g(i)g(j).$$

If  $i + j \geq m$ , then

$$g(i \oplus j) = g(i + j - m) = x^{i+j-m} = x^i x^j (x^m)^{-1} = x^i x^j = g(i)g(j).$$

So  $g$  is an isomorphism from  $(\mathbb{Z}_m, \oplus)$  to the cyclic subgroup generated by  $x$ .

We have just shown that *every cyclic group  $\langle x \rangle$  is isomorphic to one of the additive groups  $\mathbb{Z}$  or  $\mathbb{Z}_m$  for some  $m > 0$* . The first case occurs when  $x$  has infinite order, and the second case occurs when  $x$  has order  $m$ .

## 9.11 Group Actions

The fundamental tool needed to solve counting problems involving symmetry is the notion of a *group action*.

**9.80. Definition: Group Actions.** Suppose  $G$  is a group and  $X$  is a set. An *action* of  $G$  on  $X$  is a function  $* : G \times X \rightarrow X$  satisfying the following axioms.

1. For all  $g \in G$  and all  $x \in X$ ,  $g * x \in X$  (closure).
2. For all  $x \in X$ ,  $e_G * x = x$  (identity).
3. For all  $g, h \in G$  and all  $x \in X$ ,  $g * (h * x) = (gh) * x$  (associativity).

The pair  $(X, *)$  is called a  *$G$ -set*.

**9.81. Example.** For any set  $X$ , the group  $G = (\text{Sym}(X), \circ)$  acts on  $X$  via the rule  $g * x = g(x)$  for  $g \in G$  and  $x \in X$ . Axiom 1 holds because each  $g \in G$  is a function from  $X$  to  $X$ , hence  $g * x = g(x) \in X$  for all  $x \in X$ . Axiom 2 holds since  $e_G * x = \text{id}_X * x = \text{id}_X(x) = x$  for all  $x \in X$ . Axiom 3 holds because

$$g * (h * x) = g(h(x)) = (g \circ h)(x) = (gh) * x \quad (g, h \in \text{Sym}(X), x \in X).$$

**9.82. Example.** Let  $G$  be any group, written multiplicatively, and let  $X$  be the set  $G$ . Define  $* : G \times X \rightarrow X$  by  $g * x = gx$  for all  $g, x \in G$ . We say that “ $G$  acts on itself by left multiplication.” In this example, the action axioms reduce to the corresponding group axioms for  $G$ .

We can define another action of  $G$  on  $X = G$  by letting  $g \bullet x = xg^{-1}$  for all  $g, x \in G$ . The first two axioms for an action are immediately verified; the third axiom follows from the calculation

$$g \bullet (h \bullet x) = g \bullet (xh^{-1}) = (xh^{-1})g^{-1} = x(h^{-1}g^{-1}) = x(gh)^{-1} = (gh) \bullet x \quad (g, h, x \in G).$$

We say that “ $G$  acts on itself by inverted right multiplication.” One can check that the rule  $g \cdot x = xg$  (for  $g, x \in G$ ) does *not* define a group action for non-commutative groups  $G$ , because axiom 3 fails. (But see the discussion of right group actions below.)

**9.83. Example.** Let the group  $G$  act on the set  $X = G$  as follows:

$$g * x = xgx^{-1} \quad (g \in G, x \in X).$$

We say that “ $G$  acts on itself by conjugation.” The reader should verify that the axioms for an action are satisfied.

**9.84. Example.** Suppose we are given a group action  $*$  :  $G \times X \rightarrow X$ . Let  $H$  be any subgroup of  $G$ . By restricting the action function to  $H \times X$ , we obtain an action of  $H$  on  $X$ , as one immediately verifies. Combining this construction with previous examples, we obtain quite a few additional instances of group actions. For example, any subgroup  $H$  of a group  $G$  acts on  $G$  by left multiplication, and by inverted right multiplication, and by conjugation. Any subgroup  $H$  of  $\text{Sym}(X)$  acts on  $X$  via  $f \star x = f(x)$  for  $f \in H$  and  $x \in X$ . In particular, the automorphism group  $\text{Aut}(G)$  of a group  $G$  is a subgroup of  $\text{Sym}(G)$ , so  $\text{Aut}(G)$  acts on  $G$  via  $f \star x = f(x)$  for  $f \in \text{Aut}(G)$  and  $x \in G$ . Similarly, if  $K$  is a graph with vertex set  $X$ , then  $\text{Aut}(K)$  is a subgroup of  $\text{Sym}(X)$ , and therefore  $\text{Aut}(K)$  acts on  $X$  via  $f \star x = f(x)$  for  $f \in \text{Aut}(K)$  and  $x \in X$ .

**9.85. Example.** Suppose  $(X, *)$  is a  $G$ -set. Let  $\mathcal{P}(X)$  be the power set of  $X$ , which consists of all subsets of  $X$ . It is routine to check that  $\mathcal{P}(X)$  is a  $G$ -set under the action

$$g \bullet S = \{g * s : s \in S\} \quad (g \in G, S \in \mathcal{P}(X)).$$

**9.86. Example.** Consider a polynomial ring  $R = F[x_1, x_2, \dots, x_n]$ , where  $F$  is a field (see §7.16). The symmetric group  $S_n$  acts on  $\{1, 2, \dots, n\}$  via  $f * i = f(i)$  for  $f \in S_n$  and  $1 \leq i \leq n$ . We can transfer this to an action of  $S_n$  on  $\{x_1, \dots, x_n\}$  by defining

$$f * x_i = x_{f(i)} \quad (f \in S_n, 1 \leq i \leq n).$$

Using the universal mapping property of polynomial rings (see 7.102), each bijection  $(x_i \mapsto x_{f(i)} : 1 \leq i \leq n)$  extends to a ring isomorphism  $\bar{f}$  sending  $p = p(x_1, \dots, x_n) \in R$  to  $\bar{f}(p) = p(x_{f(1)}, \dots, x_{f(n)})$ . One may check that the rule  $f * p = \bar{f}(p)$  (for  $f \in S_n$  and  $p \in R$ ) defines an action of  $S_n$  on  $R$ . In particular,  $g * (h * p) = (g \circ h) * p$  follows by the uniqueness part of the universal mapping property, since both sides are the image of  $p$  under the unique ring homomorphism sending  $x_i$  to  $x_{g(h(i))}$  for all  $i$ .

**9.87. Example.** By imitating ideas in the previous example, we can define certain group actions on vector spaces. Suppose  $V$  is a vector space over a field  $F$  and let  $X = (x_1, \dots, x_n)$  be an ordered basis of  $V$ . For  $f \in S_n$ , the map  $x_i \mapsto x_{f(i)}$  on basis vectors extends by linearity to a unique linear map  $T_f : V \rightarrow V$ , given explicitly by

$$T_f(a_1x_1 + \dots + a_nx_n) = a_1x_{f(1)} + \dots + a_nx_{f(n)} \quad (a_i \in F).$$

One may check that  $f * v = T_f(v)$  (for  $f \in S_n$  and  $v \in V$ ) defines an action of the group  $S_n$  on the set  $V$ .

**9.88. Example.** Suppose  $G$  is a group,  $(X, *)$  is a  $G$ -set, and  $W$  and  $Z$  are any sets. Recall that  ${}^W X$  is the set of all functions  $F : W \rightarrow X$ . This set of functions can be turned into a  $G$ -set by defining

$$(g \bullet F)(w) = g * (F(w)) \quad (g \in G, F \in {}^W X, w \in W).$$

We leave the verification of the action axioms as an exercise.

Now consider the set  ${}^X Z$  of all functions  $F : X \rightarrow Z$ . We claim this set of functions becomes a  $G$ -set if we define

$$(g \bullet F)(x) = F(g^{-1} * x) \quad (g \in G, F \in {}^X Z, x \in X).$$

Let us carefully prove this claim. First, given  $g \in G$  and  $F \in {}^X Z$ , the map  $g \bullet F$  is a well-defined function from  $X$  to  $Z$  because  $g^{-1} * x \in X$  and  $F$  maps  $X$  into  $Z$ . So,  $g \bullet F \in {}^X Z$ , verifying closure. Second, letting  $e$  be the identity of  $G$  and letting  $F \in {}^X Z$ , we have

$$(e \bullet F)(x) = F(e^{-1} * x) = F(e * x) = F(x) \quad (x \in X),$$



so that we have the equality of functions  $e \bullet F = F$ . (Recall that two functions are equal iff they have the same domain  $X$ , have the same codomain, and take the same values at each  $x \in X$ .) Third, we verify the associativity axiom for  $\bullet$ . Fix  $g, h \in G$  and  $F \in {}^X Z$ . The two functions  $g \bullet (h \bullet F)$  and  $(gh) \bullet F$  both have domain  $X$  and codomain  $Z$ . Fix  $x \in X$ . On one hand,

$$[(gh) \bullet F](x) = F((gh)^{-1} * x) = F((h^{-1}g^{-1}) * x).$$

On the other hand, using the definition of  $\bullet$  twice,

$$[g \bullet (h \bullet F)](x) = [h \bullet F](g^{-1} * x) = F(h^{-1} * (g^{-1} * x)).$$

Since  $*$  is known to be an action, we see that  $g \bullet (h \bullet F)$  and  $(gh) \bullet F$  take the same value at  $x$ . So the third action axiom is proved. The reader may check that this axiom would fail, in general, if we omitted the inverse in the definition of  $\bullet$ .

**9.89. Example.** Let  $n$  be a fixed integer, let  $Y$  be a set, and let

$$U = \{(y_1, \dots, y_n) : y_i \in Y\}$$

be the set of all sequences of  $n$  elements of  $Y$ . The group  $S_n$  acts on  $U$  via the rule

$$f \cdot (y_1, y_2, \dots, y_n) = (y_{f^{-1}(1)}, y_{f^{-1}(2)}, \dots, y_{f^{-1}(n)}) \quad (f \in S_n; y_1, \dots, y_n \in Y).$$

The inverses in this formula are *essential*. To see why, we observe that the action here is actually a special case of the previous example. For, a sequence in  $U$  is officially defined to be a function  $y : \{1, 2, \dots, n\} \rightarrow Y$  where  $y(i) = y_i$ . Using this function notation for sequences, we have (for  $f \in S_n$ )

$$(f \cdot y)(i) = y(f^{-1}(i)) = (f \bullet y)(i) \quad (1 \leq i \leq n),$$

in agreement with the previous example. One should also note that acting by  $f$  moves the object  $z$  originally in *position*  $i$  to *position*  $f(i)$  in the new sequence. This is true because  $(f \cdot y)(f(i)) = y(f^{-1}(f(i))) = y(i) = z$ .

The reader may now be disturbed by the *lack* of inverses in the formula  $f * x_i = x_{f(i)}$  from 9.87. However, there is no contradiction since the  $x_i$ 's in the latter example are fixed basis elements in a vector space  $V$ , not the entries in a sequence. Indeed, recall that the action on  $V$  is given by  $f * v = T_f(v)$  where  $T_f$  is the linear extension of the map  $x_i \mapsto x_{f(i)}$ . Writing  $v = \sum_i a_i x_i$ , the *coordinates* of  $v$  relative to this basis are the entries in the sequence  $(a_1, a_2, \dots, a_n)$ . Applying  $f$  to  $v$  gives

$$\sum_i a_i x_{f(i)} = \sum_j a_{f^{-1}(j)} x_j,$$

where we changed variables by letting  $j = f(i)$ ,  $i = f^{-1}(j)$ . We now see that the coordinates of  $f * v$  relative to the ordered basis  $(x_1, \dots, x_n)$  are  $(a_{f^{-1}(1)}, \dots, a_{f^{-1}(n)})$ . For example,

$$(1, 2, 3) * (a_1 x_1 + a_2 x_2 + a_3 x_3) = (a_1 x_2 + a_2 x_3 + a_3 x_1) = (a_3 x_1 + a_1 x_2 + a_2 x_3),$$

or equivalently, in coordinate notation,

$$(1, 2, 3) * (a_1, a_2, a_3) = (a_3, a_1, a_2).$$

To summarize, when  $f$  acts directly on the *objects*  $x_i$ , no inverse is needed; but when  $f$  permutes the *positions* in a list, one must apply  $f^{-1}$  to each subscript.

**9.90. Remark: Right Actions.** A *right action* of a group  $G$  on a set  $X$  is a map  $*$  :  $X \times G \rightarrow X$  such that  $x * e = x$  and  $x * (gh) = (x * g) * h$  for all  $x \in X$  and all  $g, h \in G$ . For example,  $x * g = xg$  (with no inverse) defines a right action of a group  $G$  on the set  $X = G$ . Similarly, we get a *right action* of  $S_n$  on the set of sequences in the previous example by writing

$$(y_1, \dots, y_n) * f = (y_{f(1)}, \dots, y_{f(n)}).$$

Group actions (as defined at the beginning of this section) are sometimes called *left actions* to avoid confusion with right actions. We shall mostly consider left group actions in the sequel, but right actions are occasionally more convenient to use (cf. 9.109).

## 9.12 Permutation Representations

Group actions are closely related to symmetric groups. To understand the precise nature of this relationship, we need the following definition.

**9.91. Definition: Permutation Representations.** A *permutation representation* of a group  $G$  on a set  $X$  is a group homomorphism  $\phi : G \rightarrow \text{Sym}(X)$ .

This definition seems quite different from the definition of a group action given in the last section. But we will see in this section that group actions and permutation representations are essentially the same thing. Both viewpoints turn out to be pertinent in the application of group actions to problems in combinatorics and algebra.

We first show that any group action of  $G$  on  $X$  gives rise to a permutation representation of  $G$  on  $X$  in a canonical way. The key idea appears in the next definition.

**9.92. Definition: Left Multiplication Maps.** Let  $*$  :  $G \times X \rightarrow X$  be an action of the group  $G$  on the set  $X$ . For each  $g \in G$ , *left multiplication by  $G$*  (relative to this action) is the function  $L_g : X \rightarrow X$  defined by

$$L_g(x) = g * x \quad (x \in X).$$

Note that  $L_g$  does take values in  $X$ , by the closure axiom for group actions.

**9.93. Theorem: Properties of Left Multiplication Maps.** Let  $(X, *)$  be a  $G$ -set. (a)  $L_e = \text{id}_X$ . (b) For all  $g, h \in G$ ,  $L_{gh} = L_g \circ L_h$ . (c) For all  $g \in G$ ,  $L_g \in \text{Sym}(X)$ , and  $L_g^{-1} = L_{g^{-1}}$ .

*Proof.* All functions appearing here have domain  $X$  and codomain  $X$ . So it suffices to check that the relevant functions take the same value at each  $x \in X$ . For (a),  $L_e(x) = e * x = x = \text{id}_X(x)$  by the identity axiom for group actions. For (b),  $L_{gh}(x) = (gh) * x = g * (h * x) = L_g(L_h(x)) = (L_g \circ L_h)(x)$  by the associativity axiom for group actions. Finally, using (a) and (b) with  $h = g^{-1}$  shows that  $\text{id}_X = L_g \circ L_{g^{-1}}$ . Similarly,  $\text{id}_X = L_{g^{-1}} \circ L_g$ . This means that  $L_{g^{-1}}$  is the two-sided inverse of  $L_g$ ; in particular, both of these maps must be bijections.  $\square$

Using the theorem, we can pass from a group action  $*$  to a permutation representation  $\phi$  as follows. Define  $\phi : G \rightarrow \text{Sym}(X)$  by setting  $\phi(g) = L_g \in \text{Sym}(X)$  for all  $g \in G$ . By part (b) of the theorem,

$$\phi(gh) = L_{gh} = L_g \circ L_h = \phi(g) \circ \phi(h) \quad (g, h \in G),$$

and so  $\phi$  is a group homomorphism.

**9.94. Example: Cayley's Theorem.** We have seen that any group  $G$  acts on the set  $X = G$  by left multiplication. The preceding construction produces a group homomorphism  $\phi : G \rightarrow \text{Sym}(G)$ , such that  $\phi(g) = L_g = (x \mapsto gx : x \in G)$ . We claim that  $\phi$  is injective in this situation. For, suppose  $g, h \in G$  and  $L_g = L_h$ . Applying these two functions to  $e$  (the identity of  $G$ ) gives  $g = ge = L_g(e) = L_h(e) = he = h$ . It follows that  $G$  is isomorphic (via  $\phi$ ) to the image of  $\phi$ , which is a subgroup of the symmetric group  $\text{Sym}(G)$ . We have just proved *Cayley's Theorem*, which says that *any group is isomorphic to a subgroup of some symmetric group*. If  $G$  has  $n$  elements, one can check that  $\text{Sym}(G)$  is isomorphic to  $S_n$ . So every  $n$ -element group is isomorphic to a subgroup of the specific symmetric group  $S_n$ .

**9.95. Example.** Recall that, for any set  $X$ ,  $\text{Sym}(X)$  acts on  $X$  via  $f * x = f(x)$  for  $f \in \text{Sym}(X)$  and  $x \in X$ . What is the associated permutation representation  $\phi : \text{Sym}(X) \rightarrow \text{Sym}(X)$ ? First note that for  $f \in \text{Sym}(X)$ , left multiplication by  $f$  is the map  $L_f : X \rightarrow X$  such that  $L_f(x) = f * x = f(x)$ . In other words,  $L_f = f$ , so that  $\phi(f) = L_f = f$ . This means that  $\phi$  is the identity homomorphism. More generally, whenever a subgroup  $H$  of  $\text{Sym}(X)$  acts on  $X$  in the canonical way, the corresponding permutation representation is the inclusion map of  $H$  into  $\text{Sym}(X)$ .

So far, we have seen that every group action of  $G$  on  $X$  has an associated permutation representation. We can reverse this process by starting with an arbitrary permutation representation  $\phi : G \rightarrow \text{Sym}(X)$  and building a group action, as follows. Given  $\phi$ , define  $* : G \times X \rightarrow X$  by setting  $g * x = \phi(g)(x)$  for all  $g \in G$  and  $x \in X$ . Note that  $\phi(g)$  is a function with domain  $X$ , so the expression  $\phi(g)(x)$  denotes a well-defined element of  $X$ . In particular,  $*$  satisfies the closure axiom in 9.80. Since group homomorphisms preserve identities,  $\phi(e) = \text{id}_X$ , and so  $e * x = \phi(e)(x) = \text{id}_X(x) = x$  for all  $x \in X$ . So the identity axiom holds. Finally, using the fact that  $\phi$  is a group homomorphism, we calculate

$$\begin{aligned}(gh) * x &= \phi(gh)(x) = (\phi(g) \circ \phi(h))(x) \\ &= \phi(g)(\phi(h)(x)) = g * (h * x).\end{aligned}$$

So the associativity axiom holds, completing the proof that  $*$  is a group action.

The following theorem is the formal enunciation of our earlier claim that group actions and permutation representations are “essentially the same concept.”

**9.96. Theorem: Equivalence of Group Actions and Permutation Representations.** Fix a group  $G$  and a set  $X$ . Let  $A$  be the set of all group actions of  $G$  on  $X$ , and let  $P$  be the set of all permutation representations of  $G$  on  $X$ . There are mutually inverse bijections  $F : A \rightarrow P$  and  $H : P \rightarrow A$ , given by

$$F(*) = \phi : G \rightarrow \text{Sym}(X) \text{ where } \phi(g) = L_g = (x \mapsto g * x : x \in X);$$

$$H(\phi) = * : G \times X \rightarrow X \text{ where } g * x = \phi(g)(x) \quad (g \in G, x \in X).$$

*Proof.* The discussion preceding the theorem has shown that  $F$  does map the set  $A$  into the stated codomain  $P$ , and that  $H$  does map the set  $P$  into the stated codomain  $A$ . We need only verify that  $F \circ H = \text{id}_P$  and  $H \circ F = \text{id}_A$ .

To show  $F \circ H = \text{id}_P$ , fix  $\phi \in P$ , and write  $* = H(\phi)$  and  $\psi = F(*)$ . We must confirm that  $\psi = \phi : G \rightarrow \text{Sym}(X)$ . To do this, fix  $g \in G$ , and ask whether the two functions  $\psi(g), \phi(g) : X \rightarrow X$  are equal. For each  $x \in X$ ,

$$\psi(g)(x) = L_g(x) = g * x = \phi(g)(x).$$

So  $\psi(g) = \phi(g)$  for all  $g$ , hence  $\psi = \phi$  as desired.

To show  $H \circ F = \text{id}_A$ , fix  $* \in A$ , and write  $\phi = F(*)$ ,  $\bullet = H(\phi)$ . We must confirm that  $\bullet = *$ . For this, fix  $g \in G$  and  $x \in X$ . Now compute

$$g \bullet x = \phi(g)(x) = L_g(x) = g * x. \quad \square$$

**9.97. Example.** We can use permutation representations to generate new constructions of group actions. For instance, suppose  $(X, *)$  is a  $G$ -set with associated permutation representation  $\phi : G \rightarrow \text{Sym}(X)$ . Now suppose we are given a group homomorphism  $u : K \rightarrow G$ . Composing with  $\phi$  gives a homomorphism  $\phi \circ u : K \rightarrow \text{Sym}(X)$ . This is a permutation representation of  $K$  on  $X$ , which means that  $X$  can be made into a  $K$ -set in a canonical way. Specifically, by applying the map  $H$  from the theorem, we see that the  $K$ -action on  $X$  is given by

$$k \bullet x = u(k) * x \quad (k \in K, x \in X).$$

### 9.13 Stable Subsets and Orbits

One way to gain information about a group is to study its subgroups. The analogous concept for  $G$ -sets appears in the next definition.

**9.98. Definition:  $G$ -Stable Subsets.** Let  $(X, *)$  be a  $G$ -set. A subset  $Y$  of  $X$  is called a  *$G$ -stable subset* iff  $g * y \in Y$  for all  $g \in G$  and all  $y \in Y$ .

When  $Y$  is a  $G$ -stable subset, the restriction of  $*$  to  $G \times Y$  maps into the codomain  $Y$ , by definition. Since the identity axiom and associativity axiom still hold for the restricted action, we see that  $Y$  is a  $G$ -set.

Recall that every element of a group generates a cyclic subgroup. Similarly, we can pass from an element of a  $G$ -set to a  $G$ -stable subset as follows.

**9.99. Definition: Orbits.** Suppose  $(X, *)$  is a  $G$ -set, and  $x \in X$ . The  *$G$ -orbit* of  $x$  is the set

$$Gx = G * x = \{g * x : g \in G\} \subseteq X.$$

Every orbit is a  $G$ -stable subset: for, given  $h \in G$  and  $g * x \in Gx$ , the associativity axiom gives  $h * (g * x) = (hg) * x \in Gx$ . Furthermore, by the identity axiom,  $x = e * x \in Gx$  for each  $x \in X$ .

**9.100. Example.** Let  $S_5$  act on the set  $X = \{1, 2, 3, 4, 5\}$  via  $f * x = f(x)$  for  $f \in S_5$  and  $x \in X$ . For each  $i \in X$ , the orbit  $S_5 * i = \{f(i) : f \in S_5\}$  is all of  $X$ . The reason is that for any given  $j$  in  $X$ , we can find an  $f \in S_5$  such that  $f(i) = j$ ; for instance, take  $f = (i, j)$ . On the other hand, consider the subgroup  $H = \langle (1, 3)(2, 4, 5) \rangle$  of  $S_5$ . If we let  $H$  act on  $X$  via  $f * x = f(x)$  for  $f \in H$  and  $x \in X$ , we get different orbits. One may check directly that

$$H * 1 = H * 3 = \{1, 3\}, \quad H * 2 = H * 4 = H * 5 = \{2, 4, 5\}.$$

Note that the  $H$ -orbits are precisely the connected components of the digraph representing the generator  $(1, 3)(2, 4, 5)$  of  $H$ . One can verify that this holds in general whenever a cyclic subgroup of  $S_n$  acts on  $\{1, 2, \dots, n\}$ .

Now consider the action of  $A_5$  on  $X$ . As in the case of  $S_5$ , we have  $A_5 * i = X$  for all  $i \in X$ , but for a different reason. Given  $j \in X$ , we must now find an *even* permutation sending  $i$  to  $j$ . We can use the identity permutation if  $i = j$ . Otherwise, choose two distinct elements  $k, l$  that are different from  $i$  and  $j$ , and use the permutation  $(i, j)(k, l)$ .

**9.101. Example.** Let  $S_4$  act on the set  $X$  of all 4-tuples of integers by permuting positions:

$$f * (x_1, x_2, x_3, x_4) = (x_{f^{-1}(1)}, x_{f^{-1}(2)}, x_{f^{-1}(3)}, x_{f^{-1}(4)}) \quad (f \in S_4, x_i \in \mathbb{Z}).$$

The  $S_4$ -orbit of a sequence  $x = (x_1, x_2, x_3, x_4)$  consists of all possible sequences obtainable from  $x$  by permuting the entries. For example,

$$S_4 * (5, 1, 5, 1) = \{(1, 1, 5, 5), (1, 5, 1, 5), (1, 5, 5, 1), (5, 1, 1, 5), (5, 1, 5, 1), (5, 5, 1, 1)\}.$$

As another example,  $S_4 * (3, 3, 3, 3) = \{(3, 3, 3, 3)\}$  and  $S_4 * (1, 3, 5, 7)$  is the set of all 24 permutations of this list. Now consider the cyclic subgroup  $H = \langle (1, 2, 3, 4) \rangle$  of  $S_4$ . Restricting the action turns  $X$  into an  $H$ -set. When computing orbits relative to the  $H$ -action, we are only allowed to cyclically shift the elements in each 4-tuple. So, for instance,

$$\begin{aligned} H * (5, 1, 5, 1) &= \{(5, 1, 5, 1), (1, 5, 1, 5)\}; \\ H * (1, 3, 5, 7) &= \{(1, 3, 5, 7), (3, 5, 7, 1), (5, 7, 1, 3), (7, 1, 3, 5)\}. \end{aligned}$$

As before, the orbit of a given  $x \in X$  depends heavily on which group is acting on  $X$ .

**9.102. Example.** Let a group  $G$  act on itself by left multiplication:  $g * x = gx$  for  $g, x \in G$ . For every  $x \in G$ , the orbit  $Gx$  is all of  $G$ . For, given any  $y \in G$ , we have  $(yx^{-1}) * x = y$ . In the next section, we will study what happens when a subgroup  $H$  acts on  $G$  by left (or right) multiplication.

**9.103. Example: Conjugacy Classes.** Let  $G$  be a group. We have seen that  $G$  acts on itself by conjugation:  $g * x = gxg^{-1}$  for  $g, x \in G$ . The orbit of  $x \in G$  under this action is the set

$$G * x = \{gxg^{-1} : g \in G\}.$$

This set is called the *conjugacy class of  $x$  in  $G$* . For example, when  $G = S_3$ , the conjugacy classes are

$$\begin{aligned} G * \text{id} &= \{\text{id}\}; \\ G * (1, 2) = G * (1, 3) = G * (2, 3) &= \{(1, 2), (1, 3), (2, 3)\}; \\ G * (1, 2, 3) = G * (1, 3, 2) &= \{(1, 2, 3), (1, 3, 2)\}. \end{aligned}$$

One can confirm this with the aid of the identities

$$f \circ (i, j) \circ f^{-1} = (f(i), f(j)); \quad f \circ (i, j, k) \circ f^{-1} = (f(i), f(j), f(k)) \quad (f \in S_3).$$

(The generalization of this example to any  $S_n$  is discussed in §9.16.) We observe in passing that  $G * x = \{x\}$  iff  $gxg^{-1} = x$  for all  $g \in G$  iff  $gx = xg$  for all  $g \in G$  iff  $x$  commutes with every element of  $G$ . In particular, for  $G$  commutative, every conjugacy class consists of a single element.

**9.104. Example.** Let  $B = (X, E)$  be the graph representing a  $5 \times 5$  chessboard shown in Figure 9.3. Let the graph automorphism group  $G = \text{Aut}(B)$  act on  $X$  via  $f * x = f(x)$  for  $f \in G$  and  $x \in X$ . We explicitly determined the elements of  $G$  in 9.67. We can use this calculation to find all the distinct  $G$ -orbits. They are:

$$\begin{aligned} Ga &= \{a, e, z, v\} = Ge = Gz = Gv; \\ Gb &= \{b, d, j, u, y, w, q, f\} = Gd = Gj = \dots; \\ Gc &= \{c, p, x, k\}; \\ Gg &= \{g, i, t, r\}; \\ Gh &= \{h, n, s, l\}; \\ Gm &= \{m\}. \end{aligned}$$

The reader may have noticed in these examples that distinct orbits of the  $G$ -action on  $X$  are always pairwise disjoint. We now prove that this always happens.

**9.105. Theorem: Orbit Decomposition of a  $G$ -set.** Let  $(X, *)$  be a  $G$ -set. Every element  $x \in X$  belongs to exactly one  $G$ -orbit, namely  $Gx$ . In other words, the distinct  $G$ -orbits of the action  $*$  form a set partition of  $X$ .

*Proof.* Define a relation on  $X$  by setting, for  $x, y \in X$ ,  $x \sim y$  iff  $y = g * x$  for some  $g \in G$ . This relation is reflexive on  $X$ : given  $x \in G$ , we have  $x = e * x$ , so  $x \sim x$ . This relation is symmetric: given  $x, y \in X$  with  $x \sim y$ , we know  $y = g * x$  for some  $g \in G$ . A routine calculation shows that  $x = g^{-1} * y$ , so  $y \sim x$ . This relation is transitive: given  $x, y, z \in X$  with  $x \sim y$  and  $y \sim z$ , we know  $y = g * x$  and  $z = h * y$  for some  $g, h \in G$ . So  $z = h * (g * x) = (hg) * x$ , and  $x \sim z$ . Thus we have an equivalence relation on  $X$ . Recall from the proof of 2.55 that the equivalence classes of any equivalence relation on  $X$  form a set partition of  $X$ . In this situation, the equivalence classes are precisely the  $G$ -orbits, since the equivalence class of  $x$  is

$$\{y \in X : x \sim y\} = \{y \in X : y = g * x \text{ for some } g \in G\} = Gx. \quad \square$$

**9.106. Corollary.** Every group  $G$  is the disjoint union of its conjugacy classes.

Everything we have said can be adapted to give results on right actions. In particular, if  $G$  acts on  $X$  on the right, then  $X$  is partitioned into a disjoint union of the right  $G$ -orbits

$$xG = \{x * g : g \in G\} \quad (x \in X).$$

## 9.14 Cosets

The idea of a *coset* plays a central role in group theory. Cosets arise as the orbits of a certain group action.

**9.107. Definition: Right Cosets.** Let  $G$  be a group, and let  $H$  be any subgroup of  $G$ . Let  $H$  act on  $G$  by left multiplication:  $h * x = hx$  for  $h \in H$  and  $x \in G$ . The orbit of  $x$  under this action, namely

$$Hx = \{hx : h \in H\}$$

is called the *right coset of  $x$  relative to  $H$* .

By the general theory of group actions, we know that  $G$  is the disjoint union of its right cosets.

**9.108. Example.** Let  $G = S_3$  and  $H = \{\text{id}, (1, 2)\}$ . The right cosets of  $H$  in  $G$  are

$$\begin{aligned} H \text{id} = H(1, 2) &= \{\text{id}, (1, 2)\} = H; \\ H(1, 3) = H(1, 3, 2) &= \{(1, 3), (1, 3, 2)\}; \\ H(2, 3) = H(1, 2, 3) &= \{(2, 3), (1, 2, 3)\}. \end{aligned}$$

For the subgroup  $K = \{\text{id}, (1, 2, 3), (1, 3, 2)\}$ , the right cosets are

$$K \text{id} = K \text{ and } K(1, 2) = \{(1, 2), (2, 3), (1, 3)\}.$$

Note that the subgroup itself is always a right coset, but the other right cosets are not subgroups (they do not contain the identity of  $G$ ).

By letting  $H$  act on the right, we obtain the notion of a *left coset*, which will be used frequently in the sequel.

**9.109. Definition: Left Cosets.** Let  $G$  be a group, and let  $H$  be any subgroup of  $G$ . Let  $H$  act on  $G$  by right multiplication:  $x * h = xh$  for  $h \in H$  and  $x \in G$ . The orbit of  $x$  under this action, namely

$$xH = \{xh : h \in H\}$$

is called the *left coset of  $x$  relative to  $H$* .

By 9.105,  $G$  is the disjoint union of its left cosets.

**9.110. Example.** Let  $G = S_3$  and  $H = \{\text{id}, (1, 2)\}$  as above. The left cosets of  $H$  in  $G$  are

$$\begin{aligned} \text{id}H &= (1, 2)H &= \{\text{id}, (1, 2)\} &= H; \\ (1, 3)H &= (1, 2, 3)H &= \{(1, 3), (1, 2, 3)\}; \\ (2, 3)H &= (1, 3, 2)H &= \{(2, 3), (1, 3, 2)\}. \end{aligned}$$

Observe that  $xH \neq Hx$  except when  $x \in H$ . This shows that left cosets and right cosets do not coincide in general. On the other hand, for the subgroup  $K = \{\text{id}, (1, 2, 3), (1, 3, 2)\}$ , the left cosets are  $K$  and  $(1, 2)K = \{(1, 2), (1, 3), (2, 3)\}$ . One checks that  $xK = Kx$  for all  $x \in S_3$ , so that left cosets and right cosets do coincide for some subgroups.

Although  $x = y$  certainly implies  $xH = yH$ , one must remember that the converse is almost always false. The next result gives criteria for deciding when two cosets  $xH$  and  $yH$  are equal; it is used constantly in arguments involving cosets.

**9.111. Coset Equality Theorem.** Let  $H$  be a subgroup of  $G$ . For all  $x, y \in G$ , the following conditions are logically equivalent:

---

(a) $xH = yH$ .	(a') $yH = xH$ .
(b) $x \in yH$ .	(b') $y \in xH$ .
(c) There exists $h \in H$ with $x = yh$ .	(c') There exists $h' \in H$ with $y = xh'$ .
(d) $y^{-1}x \in H$ .	(d') $x^{-1}y \in H$ .

---

*Proof.* We first prove (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (a). If  $xH = yH$ , then  $x = xe \in xH = yH$ , so  $x \in yH$ . If  $x \in yH$ , then  $x = yh$  for some  $h \in H$  by definition of  $yH$ . If  $x = yh$  for some  $h \in H$ , then multiplying by  $y^{-1}$  on the left gives  $y^{-1}x \in H$ . Finally, assume that  $y^{-1}x \in H$ . Then  $y(y^{-1}x) = x$  lies in the orbit  $yH$ . We also have  $x = xe \in xH$ . As orbits are either disjoint or equal, we must have  $xH = yH$ .

Interchanging  $x$  and  $y$  in the last paragraph proves the equivalence of (a'), (b'), (c'), and (d'). Since (a) and (a') are visibly equivalent, the proof is complete.  $\square$

**9.112. Remark.** The equivalence of (a) and (d) in the last theorem is used quite frequently. Note too that the subgroup  $H$  is a coset (namely  $eH$ ), and  $xH = H$  iff  $xH = eH$  iff  $e^{-1}x \in H$  iff  $x \in H$ . Finally, one can prove an analogous theorem for right cosets. The key difference is that  $Hx = Hy$  iff  $xy^{-1} \in H$  iff  $yx^{-1} \in H$  (so that inverses occur on the right for right cosets).

We can use cosets to construct more examples of  $G$ -sets.

**9.113. The  $G$ -set  $G/H$ .** Let  $G$  be a group, and let  $H$  be *any* subgroup of  $G$ . Let  $G/H$  be the set of all distinct left cosets of  $H$  in  $G$ . Every element of  $G/H$  is a subset of  $G$  of the

form  $xH = \{xh : h \in H\}$  for some  $x \in G$  (which is usually not unique). So,  $G/H$  is a subset of the power set  $\mathcal{P}(G)$ . Let the group  $G$  act on the set  $X = G/H$  by left multiplication:

$$g * S = \{gs : s \in S\} \quad (g \in G, S \in X).$$

Note that this action is the restriction of the action from 9.85 to  $G \times X$ . To see that the action makes sense, we must check that  $X$  is a  $G$ -stable subset of  $\mathcal{P}(G)$ . Let  $xH$  be an element of  $X$  and let  $g \in G$ ; then

$$g * (xH) = \{g(xh) : h \in H\} = \{(gx)h : h \in H\} = (gx)H \in X.$$

Let  $[G : H] = |G/H|$  (which may be infinite); this cardinal number is called the *index of  $H$  in  $G$* . Lagrange's Theorem (below) will show that  $|G/H| = |G|/|H|$  when  $G$  is finite.

**9.114. Remark.** Using the coset equality theorems, one can show that  $xH \mapsto Hx^{-1}$  gives a well-defined bijection between  $G/H$  and the set of right cosets of  $H$  in  $G$ . So, we would obtain the same number  $[G : H]$  if we had used right cosets in the definition of  $G/H$ . It is more convenient to use left cosets here, so that  $G$  can act on  $G/H$  on the left.

**9.115. Example.** If  $G = S_3$  and  $H = \{\text{id}, (1, 2)\}$ , then

$$G/H = \{\{\text{id}, (1, 2)\}, \{(1, 3), (1, 2, 3)\}, \{(2, 3), (1, 3, 2)\}\} = \{\text{id}H, (1, 3)H, (2, 3)H\}.$$

We have  $[G : H] = |G/H| = 3$ . Note that  $|G|/|H| = 6/2 = 3 = |G/H|$ . This is a special case of Lagrange's Theorem, proved below.

To prepare for Lagrange's Theorem, we first show that every left coset of  $H$  in  $G$  has the same cardinality as  $G$ .

**9.116. Coset Size Theorem.** Let  $H$  be a subgroup of  $G$ . For all  $x \in G$ ,  $|xH| = |H|$ .

*Proof.* We have seen that the left multiplication  $L_x : G \rightarrow G$ , given by  $g \mapsto xg$  for  $g \in G$ , is a bijection (with inverse  $L_{x^{-1}}$ ). Restricting the domain of  $L_x$  to  $H$  gives an injective map  $L'_x : H \rightarrow G$ . The image of this map is  $\{xh : h \in H\} = xH$ . So, restricting the codomain gives a bijection from  $H$  to  $xH$ . Thus, the sets  $H$  and  $xH$  have the same cardinality.  $\square$

**9.117. Lagrange's Theorem.** Let  $H$  be any subgroup of a finite group  $G$ . Then

$$[G : H] \cdot |H| = |G|.$$

So  $|H|$  and  $[G : H]$  are divisors of  $|G|$ , and  $|G/H| = [G : H] = |G|/|H|$ .

*Proof.* We know that  $G$  is the disjoint union of its distinct left cosets:  $G = \bigcup_{S \in G/H} S$ . By the previous theorem,  $|S| = |H|$  for every  $S \in G/H$ . So, by the sum rule,

$$|G| = \sum_{S \in G/H} |S| = \sum_{S \in G/H} |H| = |G/H| \cdot |H| = [G : H] \cdot |H|. \quad \square$$

**9.118. Remark.** The equality of cardinal numbers  $|H| \cdot [G : H] = |G|$  holds even when  $G$  is infinite, with the same proof.

**9.119. Theorem: Order of Group Elements.** If  $G$  is a finite group of size  $n$  and  $x \in G$ , then the order of  $x$  is a divisor of  $n$ , and  $x^n = e_G$ .

*Proof.* Consider the subgroup  $H = \langle x \rangle$  generated by  $x$ . The order  $d$  of  $x$  is  $|H|$ , which divides  $|G| = n$  by Lagrange's theorem. Writing  $n = cd$ , we see that  $x^n = (x^d)^c = e^c = e$ .  $\square$



The next result gives an interpretation for cosets  $xK$  in the case where  $K$  is the kernel of a group homomorphism.

**9.120. Theorem: Cosets of the Kernel of a Homomorphism.** Let  $f : G \rightarrow L$  be a group homomorphism with kernel  $K$ . For every  $x \in G$ ,

$$xK = \{y \in G : f(y) = f(x)\} = Kx.$$

If  $G$  is finite and  $I$  is the image of  $f$ , it follows that  $|G| = |K| \cdot |I|$ .

*Proof.* Fix  $x \in G$ , and set  $S = \{y \in G : f(y) = f(x)\}$ . We will prove that  $xK = S$ . First suppose  $y \in xK$ , so  $y = xk$  for some  $k \in K$ . Applying  $f$ , we find that  $f(y) = f(xk) = f(x)f(k) = f(x)e_L = f(x)$ , so  $y \in S$ . Next suppose  $y \in S$ , so  $f(y) = f(x)$ . Note that  $f(x^{-1}y) = f(x)^{-1}f(y) = e$ , so  $x^{-1}y \in \ker(f) = K$ . So  $y = x(x^{-1}y) \in xK$ . The proof that  $S = Kx$  is analogous. To obtain the formula for  $|G|$ , note that  $G$  is the disjoint union

$$G = \bigcup_{z \in I} \{y \in G : f(y) = z\}.$$

Every  $z \in I$  has the form  $z = f(x)$  for some  $x \in G$ . So, by what we have just proved, each set appearing in the union is a coset of  $K$ , which has the same cardinality as  $K$ . So the sum rule gives  $|G| = \sum_{z \in I} |K| = |K| \cdot |I|$ .  $\square$

**9.121. Corollary: Size of  $A_n$ .** For  $n > 1$ ,  $|A_n| = n!/2$ .

*Proof.* We know that  $\text{sgn} : S_n \rightarrow \{1, -1\}$  is a surjective group homomorphism with kernel  $A_n$ . So  $n! = |S_n| = |A_n| \cdot 2$ .  $\square$

## 9.15 The Size of an Orbit

In 9.105, we saw that every  $G$ -set  $X$  breaks up into a disjoint union of orbits. This result suggests two combinatorial questions. First, given  $x \in X$ , what is the size of the orbit  $Gx$ ? Second, how many orbits are there? We answer the first question here; the second question will be solved in §9.18.

The key to computing the orbit size  $|Gx|$  is to relate the  $G$ -set  $Gx$  to one of the special  $G$ -sets  $G/H$  defined in 9.113. For this purpose, we need to associate a subgroup  $H$  of  $G$  to the given orbit  $Gx$ .

**9.122. Definition: Stabilizers.** Let  $(X, *)$  be a  $G$ -set. For each  $x \in X$ , the *stabilizer of  $x$  in  $G$*  is

$$\text{Stab}(x) = \{g \in G : g * x = x\}.$$

Sometimes the notation  $G_x$  is used to denote  $\text{Stab}(x)$ .

The following calculations show that  $\text{Stab}(x)$  is a *subgroup* of  $G$  for each  $x \in X$ :  $e * x = x$ ;  $g * x = x$  implies  $x = g^{-1} * x$  for  $g \in G$ ;  $g * x = x = h * x$  implies  $(gh) * x = g * (h * x) = x$  for  $g, h \in G$ .

**9.123. Example.** Let  $S_n$  act on  $X = \{1, 2, \dots, n\}$  via  $f * x = f(x)$  for  $f \in S_n$  and  $x \in X$ . The stabilizer of a point  $i \in X$  consists of all permutations of  $X$  for which  $i$  is a fixed point. In particular,  $\text{Stab}(n)$  consist of all bijections  $f : X \rightarrow X$  with  $f(n) = n$ . Restricting the domain to  $\{1, 2, \dots, n-1\}$  defines a group isomorphism between  $\text{Stab}(n)$  and  $S_{n-1}$ .

**9.124. Example.** Let a group  $G$  act on itself by left multiplication. Right cancellation of  $x$  shows that  $gx = x$  iff  $g = e$ . Therefore,  $\text{Stab}(x) = \{e\}$  for all  $x \in G$ . At the other extreme, we can let  $G$  act on any set  $X$  by declaring  $g * x = x$  for all  $g \in G$  and all  $x \in X$ . Relative to this action,  $\text{Stab}(x) = G$  for all  $x \in X$ .

**9.125. Example: Centralizers.** Let  $G$  act on itself by conjugation:  $g * x = gxg^{-1}$  for all  $g, x \in G$ . For a given  $x \in G$ ,  $g \in \text{Stab}(x)$  iff  $gxg^{-1} = x$  iff  $gx = xg$  iff  $g$  commutes with  $x$ . This stabilizer subgroup is often denoted  $C_G(x)$  and called the *centralizer of  $x$  in  $G$* . The intersection  $\bigcap_{x \in G} C_G(x)$  consists of all  $g \in G$  that commute with *every*  $x \in G$ . This is a subgroup called the *center* of  $G$  and denoted  $Z(G)$ .

**9.126. Example: Normalizers.** Let  $G$  be a group, and let  $X$  be the set of all subgroups of  $G$ .  $G$  acts on  $X$  by conjugation:  $g * H = gHg^{-1} = \{ghg^{-1} : h \in H\}$ . (Note that  $g * H$  is a subgroup, since it is the image of a subgroup under the inner automorphism “conjugation by  $g$ ”; cf. 9.74.) For this action,  $g \in \text{Stab}(H)$  iff  $gHg^{-1} = H$ . This stabilizer subgroup is denoted  $N_G(H)$  and called the *normalizer of  $H$  in  $G$* . One may check that  $N_G(H)$  always contains  $H$ .

**9.127. Example.** Let  $S_4$  act on 4-tuples of integers by permuting the positions. Then  $\text{Stab}((5, 1, 5, 1)) = \{\text{id}, (1, 3), (2, 4), (1, 3)(2, 4)\}$ ;  $\text{Stab}((2, 2, 2, 2)) = S_4$ ;  $\text{Stab}((1, 2, 3, 4)) = \{\text{id}\}$ ; and  $\text{Stab}((2, 5, 2, 2))$  is a subgroup of  $S_4$  isomorphic to  $\text{Sym}(\{1, 3, 4\})$ , which is in turn isomorphic to  $S_3$ .

The following fundamental theorem calculates the size of an orbit of a group action.

**9.128. Theorem: Size of an Orbit.** Let  $(X, *)$  be a  $G$ -set. For each  $x \in X$ , there is a bijection  $f : G/\text{Stab}(x) \rightarrow Gx$  given by  $f(g\text{Stab}(x)) = g * x$  for all  $g \in G$ . So, when  $G$  is finite, *the size of the orbit of  $x$  is the index of the stabilizer of  $x$ , which is a divisor of  $|G|$ :*

$$|Gx| = [G : \text{Stab}(x)] = |G|/|\text{Stab}(x)|.$$

*Proof.* Write  $H = \text{Stab}(x)$  for convenience. We first check that the function  $f : G/H \rightarrow Gx$  is well defined. Assume  $g, k \in G$  satisfy  $gH = kH$ ; we must check that  $g * x = k * x$ . Now,  $gH = kH$  means  $k^{-1}g \in H = \text{Stab}(x)$ , and hence  $(k^{-1}g) * x = x$ . Acting on both sides by  $k$  and simplifying, we obtain  $g * x = k * x$ . Second, is  $f$  one-to-one? Fix  $g, k \in G$  with  $f(gH) = f(kH)$ ; we must prove  $gH = kH$ . Now,  $f(gH) = f(kH)$  means  $g * x = k * x$ . Acting on both sides by  $k^{-1}$ , we find that  $(k^{-1}g) * x = x$ , so  $k^{-1}g \in H$ , so  $gH = kH$ . Third, is  $f$  surjective? Given  $y \in Gx$ , the definition of  $Gx$  says that  $y = g * x$  for some  $g \in G$ , so  $y = f(gH)$ . In summary,  $f$  is a well-defined bijection.  $\square$

**9.129. Remark.** One can prove a stronger version of the theorem, analogous to the “fundamental homomorphism theorem for groups,” by introducing the following definition. Given two  $G$ -sets  $(X, *)$  and  $(Y, \bullet)$ , a  $G$ -map is a function  $p : X \rightarrow Y$  such that  $p(g * x) = g \bullet p(x)$  for all  $g \in G$  and all  $x \in X$ . A  $G$ -isomorphism is a bijective  $G$ -map. The theorem gives us a bijection  $p$  from the  $G$ -set  $Gx$  to the  $G$ -set  $G/\text{Stab}(x)$  such that  $p(g_0 * x) = g_0 \text{Stab}(x)$ . This bijection is in fact a  $G$ -isomorphism, because

$$p(g * (g_0 * x)) = p((gg_0) * x) = (gg_0) \text{Stab}(x) = g \bullet (g_0 \text{Stab}(x)) = g \bullet p(g_0 * x).$$

Since every  $G$ -set is a disjoint union of orbits, this result shows that the special  $G$ -sets of the form  $G/H$  are the “building blocks” from which all  $G$ -sets are constructed.

Applying 9.128 to some of the preceding examples gives the following corollary.

**9.130. Corollary: Counting Conjugates of Group Elements and Subgroups.** The size of the conjugacy class of  $x$  in a finite group  $G$  is  $[G : \text{Stab}(x)] = [G : C_G(x)] = |G|/|C_G(x)|$ . If  $H$  is a subgroup of  $G$ , the number of distinct conjugates of  $H$  (subgroups of the form  $gHg^{-1}$ ) is  $[G : \text{Stab}(H)] = [G : N_G(H)] = |G|/|N_G(H)|$ .

## 9.16 Conjugacy Classes in $S_n$

The conjugacy classes in the symmetric groups  $S_n$  can be described explicitly. We shall prove that the conjugacy class of  $f \in S_n$  consists of all  $g \in S_n$  with the same cycle type as  $f$  (see 9.20). The proof employs the following computational result.

**9.131. Theorem: Conjugation in  $S_n$ .** For  $f, g \in S_n$ , the permutation  $gfg^{-1}$  can be obtained by applying  $g$  to each entry in the disjoint cycle decomposition of  $f$ . In other words, if

$$f = (i_1, i_2, i_3, \dots)(j_1, j_2, \dots)(k_1, k_2, \dots) \cdots,$$

then

$$gfg^{-1} = (g(i_1), g(i_2), g(i_3), \dots)(g(j_1), g(j_2), \dots)(g(k_1), g(k_2), \dots) \cdots.$$

In particular,  $\text{type}(gfg^{-1}) = \text{type}(f)$ .

*Proof.* First assume  $f$  is a  $k$ -cycle, say  $f = (i_1, i_2, \dots, i_k)$ . We prove that the functions  $gfg^{-1}$  and  $h = (g(i_1), g(i_2), \dots, g(i_k))$  are equal by showing that both have the same effect on every  $x \in \{1, 2, \dots, n\}$ . We consider various cases. First, if  $x = g(i_s)$  for some  $s < k$ , then  $gfg^{-1}(x) = gfg^{-1}(g(i_s)) = g(f(i_s)) = g(i_{s+1}) = h(x)$ . Second, if  $x = g(i_k)$ , then  $gfg^{-1}(x) = g(f(i_k)) = g(i_1) = h(x)$ . Finally, if  $x$  does not equal any  $g(i_s)$ , then  $g^{-1}(x)$  does not equal any  $i_s$ . So  $f$  fixes  $g^{-1}(x)$ , and  $gfg^{-1}(x) = g(g^{-1}(x)) = x = h(x)$ .

In the general case, write  $f = C_1 \circ C_2 \circ \cdots \circ C_t$  where each  $C_i$  is a cycle. Since conjugation by  $g$  is a homomorphism,

$$gfg^{-1} = (gC_1g^{-1}) \circ (gC_2g^{-1}) \circ \cdots \circ (gC_tg^{-1}).$$

By the previous paragraph, we can compute  $gC_i g^{-1}$  by applying  $g$  to each element of  $C_i$ . This completes the proof.  $\square$

**9.132. Theorem: Conjugacy Classes of  $S_n$ .** The conjugacy class of  $f \in S_n$  consists of all  $h \in S_n$  with  $\text{type}(h) = \text{type}(f)$ . The number of conjugacy classes is  $p(n)$ , the number of integer partitions of  $n$ .

*Proof.* Fix  $f \in S_n$ ; let  $T = \{gfg^{-1} : g \in S_n\}$  be the conjugacy class of  $f$ , and let  $U = \{h \in S_n : \text{type}(h) = \text{type}(f)\}$ . Using 9.131, we see that  $T \subseteq U$ . For the reverse inclusion, let  $h \in S_n$  have the same cycle type of  $f$ . We give an algorithm for finding a  $g \in S_n$  such that  $h = gfg^{-1}$ . Write down any complete cycle decomposition of  $f$  (including 1-cycles), writing longer cycles before shorter cycles. Immediately below this, write down a complete cycle decomposition of  $h$ . Now erase all the parentheses and regard the resulting array as the two-line form of a permutation  $g$ . Then 9.131 shows that  $gfg^{-1} = h$ . For example, suppose

$$\begin{aligned} f &= (1, 7, 3)(2, 8, 9)(4, 5)(6) \\ h &= (4, 9, 2)(6, 3, 5)(1, 8)(7) \end{aligned}$$

Then

$$g = \begin{pmatrix} 1 & 7 & 3 & 2 & 8 & 9 & 4 & 5 & 6 \\ 4 & 9 & 2 & 6 & 3 & 5 & 1 & 8 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 6 & 2 & 1 & 8 & 7 & 9 & 3 & 5 \end{pmatrix}.$$

The  $g$  constructed here is not unique; we could obtain different  $g$ 's satisfying  $gfg^{-1} = h$  by starting with a different complete cycle decomposition for  $f$  or  $h$ .

The last statement of the theorem follows since the possible cycle types of permutations of  $n$  objects are exactly the integer partitions of  $n$  (weakly decreasing sequences of positive integers that sum to  $n$ ).  $\square$

We now apply 9.130 to determine the sizes of the conjugacy classes of  $S_n$ .

**9.133. Definition:**  $z_\mu$ . Let  $\mu$  be an integer partition of  $n$  consisting of  $a_1$  ones,  $a_2$  twos, etc. Define

$$z_\mu = 1^{a_1} 2^{a_2} \cdots n^{a_n} a_1! a_2! \cdots a_n!.$$

For example, for  $\mu = (3, 3, 2, 2, 2, 1, 1, 1, 1)$ , we have  $a_1 = 5$ ,  $a_2 = 4$ ,  $a_3 = 2$ , and  $z_\mu = 1^5 2^4 3^2 5! 4! 2! = 829,440$ .

**9.134. Theorem: Size of Conjugacy Classes of  $S_n$ .** For each  $\mu \in \text{Par}(n)$ , the number of permutations  $f \in S_n$  with  $\text{type}(f) = \mu$  is  $n!/z_\mu$ .

*Proof.* Fix a particular  $f \in S_n$  with  $\text{type}(f) = \mu$ . By 9.130 and the fact that  $|S_n| = n!$ , it is enough to show that  $|C_{S_n}(f)| = z_\mu$ . The argument is most readily understood by consideration of a specific example. Let  $\mu = (3, 3, 2, 2, 2, 1, 1, 1, 1)$  as above, and take

$$f = (1, 2, 3)(4, 5, 6)(7, 8)(9, 10)(11, 12)(13, 14)(15)(16)(17)(18)(19).$$

A permutation  $g \in S_n$  lies in  $C_{S_n}(f)$  iff  $gfg^{-1} = f$  iff applying  $g$  to each symbol in the cycle decomposition above produces another cycle decomposition of  $f$ . So we are reduced to counting the number of ways of writing down a complete cycle decomposition of  $f$  such that longer cycles come before shorter cycles. Note that we have freedom to rearrange the order of all cycles of a given length, and we also have freedom to cyclically permute the entries in any given cycle of  $f$ . For example, we could permute the five 1-cycles of  $f$  in any of  $5!$  ways; we could replace  $(4, 5, 6)$  by one of the three cyclic shifts  $(4, 5, 6)$  or  $(5, 6, 4)$  or  $(6, 4, 5)$ ; and so on. For this particular  $f$ , the product rule gives  $2!4!5!3^2 2^4 1^5 = z_\mu$  different possible complete cycle decompositions. The argument for the general case is similar: the term  $a_i!$  in  $z_\mu$  accounts for permuting the  $a_i$  cycles of length  $i$ , while the term  $i^{a_i}$  accounts for the  $i$  possible cyclic shifts of each of the  $a_i$  cycles of length  $i$ . Multiplying these contributions gives  $z_\mu$ , as desired.  $\square$

## 9.17 Applications of the Orbit Size Formula

When a finite group  $G$  acts on a finite set  $X$ , 9.128 asserts that the size of the orbit  $Gx$  is  $|G|/|\text{Stab}(x)|$ , which is a divisor of  $|G|$ . We now use this fact to establish several famous theorems from algebra, number theory, and combinatorics.

**9.135. Fermat's Little Theorem.** For every integer  $a > 0$  and every prime  $p$ ,  $a^p \equiv a \pmod{p}$ .

*Proof.* Let  $Y = \{1, 2, \dots, a\}$ , and let  $X = Y^p$  be the set of all  $p$ -tuples  $(y_1, \dots, y_p)$  of elements of  $Y$ . By the product rule,  $|X| = a^p$ . We know that  $S_p$  acts on  $X$  by permuting positions (see 9.89). Let  $H = \langle (1, 2, \dots, p) \rangle$ , which is a cyclic subgroup of  $S_p$  of size  $p$ . Restricting the action to  $H$ , we see that  $H$  acts on  $X$  by cyclically shifting positions. The only divisors of the prime  $p$  are 1 and  $p$ , so all orbits of  $X$  under the  $H$ -action have size 1 or  $p$ . Since  $X$  is the disjoint union of the orbits,  $|X|$  is congruent modulo  $p$  to the number of orbits of size 1. But one sees immediately that  $w = (y_1, \dots, y_p)$  is in an orbit of size 1 iff all cyclic shifts of  $w$  are equal to  $w$  iff  $y_1 = \dots = y_p \in Y$ . So there are precisely  $a$  orbits of size 1, as desired.  $\square$

**9.136. Cauchy's Theorem.** Suppose  $G$  is a finite group and  $p$  is a prime divisor of  $|G|$ . Then there exists an element  $x \in G$  of order  $p$ .

*Proof.* As in the previous proof, the group  $H = \langle (1, 2, \dots, p) \rangle$  acts on the set  $G^p$  by cyclically permuting positions. Let  $X$  consist of all  $p$ -tuples  $(g_1, \dots, g_p) \in G^p$  such that  $g_1 g_2 \cdots g_p = e$ . We can build a typical element of  $X$  by choosing  $g_1, \dots, g_{p-1}$  arbitrarily from  $G$ ; then we are forced to choose  $g_p = (g_1 \cdots g_{p-1})^{-1}$  to achieve the condition  $g_1 g_2 \cdots g_{p-1} g_p = e$ . The product rule therefore gives  $|X| = |G|^{p-1}$ , which is a multiple of  $p$ .

We next claim that  $X$  is an  $H$ -stable subset of  $G^p$ . This means that for all  $i \leq p$ ,  $g_1 g_2 \cdots g_p = e$  implies  $g_i g_{i+1} \cdots g_p g_1 \cdots g_{i-1} = e$ . To prove this, multiply the equation  $g_1 g_2 \cdots g_p = e$  by  $(g_1 g_2 \cdots g_{i-1})^{-1}$  on the left and by  $(g_1 g_2 \cdots g_{i-1})$  on the right. We now know that  $X$  is an  $H$ -set, so it is a union of orbits of size 1 and size  $p$ . Since  $|X|$  is a multiple of  $p$ , the number of orbits of size 1 must be a multiple of  $p$  as well. Now,  $(e, e, \dots, e)$  is one orbit of size 1; so there must exist at least  $p - 1 > 0$  additional orbits of size 1. By definition of the  $H$ -action, such an orbit looks like  $(x, x, \dots, x)$  where  $x \neq e$ . By definition of  $X$ , we must have  $x^p = e$ . Since  $p$  is prime, we have proved the existence of an element  $x$  of order  $p$  (in fact, we know there are at least  $p - 1$  such elements).  $\square$

**9.137. Lucas' Congruence for Binomial Coefficients.** Suppose  $p$  is prime and  $0 \leq k \leq n$  are integers. Let  $n$  and  $k$  have base- $p$  expansions  $n = \sum_{i \geq 0} n_i p^i$ ,  $k = \sum_{i \geq 0} k_i p^i$ , where  $0 \leq n_i, k_i < p$  (see 5.5). Then

$$\binom{n}{k} \equiv \prod_{i \geq 0} \binom{n_i}{k_i} \pmod{p}, \quad (9.6)$$

where we set  $\binom{0}{0} = 1$  and  $\binom{a}{b} = 0$  whenever  $b > a$ .

*Proof. Step 1:* For all  $j \geq 0$ ,  $m \geq 0$ , and  $p$  prime, we show that

$$\binom{m+p}{j} \equiv \binom{m}{j} + \binom{m}{j-p} \pmod{p}. \quad (9.7)$$

To prove this identity, let  $X = \{1, 2, \dots, m+p\}$ , and let  $Y$  be the set of all  $j$ -element subsets of  $X$ . We know that  $|Y| = \binom{m+p}{j}$ . Consider the subgroup  $G = \langle (1, 2, \dots, p) \rangle$  of  $\text{Sym}(X)$ , which is cyclic of size  $p$ .  $G$  acts on  $Y$  via  $g \star S = \{g(s) : s \in S\}$  for  $g \in G$  and  $S \in Y$ .

$Y$  is a disjoint union of orbits under this action. Since every orbit has size 1 or  $p$ ,  $|Y|$  is congruent modulo  $p$  to the number  $M$  of orbits of size 1. We will show that  $M = \binom{m}{j} + \binom{m}{j-p}$ . The orbits of size 1 correspond to the  $j$ -element subsets  $S$  of  $X$  such that  $g \star S = S$  for all  $g \in G$ . It is equivalent to require that  $f \star S = S$  for the generator  $f = (1, 2, \dots, p)$  of  $G$ . Suppose  $S$  satisfies this condition, and consider two cases. Case 1:  $S \cap \{1, 2, \dots, p\} = \emptyset$ . Since  $f(x) = x$  for  $x > p$ , we have  $f \star S = S$  for all such subsets  $S$ . Since  $S$  can be an arbitrary subset of the  $m$ -element set  $\{p+1, \dots, m+p\}$ , there are  $\binom{m}{j}$  subsets of this form. Case 2:  $S \cap \{1, 2, \dots, p\} \neq \emptyset$ . Say  $i \in S$  where  $1 \leq i \leq p$ . Applying  $f$  repeatedly and noting that  $f \star S = S$ , we see that  $\{1, 2, \dots, p\} \subseteq S$ . The remaining  $j-p$  elements of  $S$  can be chosen arbitrarily from the  $m$ -element set  $\{p+1, \dots, m+p\}$ . So there are  $\binom{m}{j-p}$  subsets of this form. Combining the two cases, we see that  $M = \binom{m}{j} + \binom{m}{j-p}$ .

*Step 2:* Assume  $p$  is prime,  $a, c \geq 0$ , and  $0 \leq b, d < p$ ; we show that  $\binom{ap+b}{cp+d} \equiv \binom{a}{c} \binom{b}{d} \pmod{p}$ . This will follow from step 1 and the identity  $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$  (see 2.25). We argue by induction on  $a$ . The base step is  $a = 0$ . If  $a = 0$  and  $c > 0$ , both sides of the congruence are zero; if  $a = 0 = c$ , then both sides of the congruence are  $\binom{b}{d}$ . Assuming that the result holds for a given  $a$  (and all  $b, c, d$ ), the following computation shows that it holds

for  $a + 1$ :

$$\begin{aligned} \binom{(a+1)p+b}{cp+d} &= \binom{(ap+b)+p}{cp+d} \equiv \binom{ap+b}{cp+d} + \binom{ap+b}{(c-1)p+d} \equiv \binom{a}{c} \binom{b}{d} + \binom{a}{c-1} \binom{b}{d} \\ &= \left[ \binom{a}{c} + \binom{a}{c-1} \right] \binom{b}{d} = \binom{a+1}{c} \binom{b}{d} \pmod{p}. \end{aligned}$$

*Step 3:* We prove Lucas' congruence (9.6) by induction on  $n$ . If  $k > n$ , then  $k_i > n_i$  for some  $i$ , so that both sides of the congruence are zero. From now on, assume  $k \leq n$ . The result holds in the base cases  $0 \leq n < p$ , since  $n = n_0$ ,  $k = k_0$ , and all higher digits of the base  $p$  expansion are zero. For the induction step, note that  $n = ap + n_0$ ,  $k = cp + k_0$ , where  $a = \sum_{i \geq 0} n_{i+1}p^i$  and  $c = \sum_{i \geq 0} k_{i+1}p^i$  in base  $p$ . (We obtain  $a$  and  $c$  from  $n$  and  $k$ , respectively, by chopping off the final base  $p$  digits  $n_0$  and  $k_0$ .) By step 2 and induction, we have

$$\binom{n}{p} \equiv \binom{a}{c} \binom{n_0}{k_0} \equiv \binom{n_0}{k_0} \prod_{i \geq 1} \binom{n_i}{k_i} = \prod_{i \geq 0} \binom{n_i}{k_i} \pmod{p}. \quad \square$$

**9.138. Corollary.** Given  $a, b, p \in \mathbb{N}^+$  with  $p$  prime and  $p$  not dividing  $b$ ,

$$p \text{ does not divide } \binom{p^a b}{p^a}.$$

*Proof.* Write  $b = \sum_{i \geq 0} b_i p^i$  in base  $p$ . The base- $p$  expansions of  $p^a b$  and  $p^a$  are  $p^a b = \dots b_3 b_2 b_1 b_0 00 \dots 0$  and  $p^a = 1000 \dots 0$ , respectively, where each expansion ends in  $a$  zeroes. Since  $b_0 \neq 0$  by hypothesis, Lucas' congruence gives

$$\binom{p^a b}{p^a} \equiv \binom{b_0}{1} = b_0 \not\equiv 0 \pmod{p}. \quad \square$$

This corollary can also be proved directly, by writing out the fraction defining  $\binom{p^a b}{p^a}$  and counting powers of  $p$  in numerator and denominator. We leave this as an exercise for the reader.

**9.139. Sylow's First Theorem.** Let  $G$  be a finite group of size  $p^a b$ , where  $p$  is prime,  $a > 0$ , and  $p$  does not divide  $b$ . There exists a subgroup  $H$  of  $G$  of size  $p^a$ .

*Proof.* Let  $X$  be the collection of all subsets of  $G$  of size  $p^a$ . We know from 1.42 that  $|X| = \binom{p^a b}{p^a}$ . By 9.138,  $p$  does not divide  $|X|$ . Now,  $G$  acts on  $X$  by left multiplication:  $g * S = \{gs : s \in S\}$  for  $g \in G$  and  $S \in X$ . (The set  $g * S$  still has size  $p^a$ , since left multiplication by  $g$  is injective.) Not every orbit of  $X$  has size divisible by  $p$ , since  $|X|$  itself is not divisible by  $p$ . Choose  $T \in X$  such that  $|GT| \not\equiv 0 \pmod{p}$ . Let  $H = \text{Stab}(T) = \{g \in G : g * T = T\}$ , which is a subgroup of  $G$ . The size of the orbit of  $T$  is  $|G|/|H| = p^a b/|H|$ . This integer is not divisible by  $p$ , forcing  $|H|$  to be a multiple of  $p^a$ . So  $|H| \geq p^a$ . To obtain the reverse inequality, let  $t_0$  be any fixed element of  $T$ . Given any  $h \in H$ ,  $h * T = T$  implies  $ht_0 \in T$ . So the right coset  $Ht_0 = \{ht_0 : h \in H\}$  is contained in  $T$ . We conclude that  $|H| = |Ht_0| \leq |T| = p^a$ . Thus  $H$  is a subgroup of size  $p^a$  (and  $T$  is in fact one of the right cosets of  $H$ ).  $\square$

## 9.18 The Number of Orbits

The following theorem, which is traditionally known as "Burnside's Lemma," allows us to count the number of orbits in a given  $G$ -set.

**9.140. Orbit-Counting Theorem.** Let a finite group  $G$  act on a finite set  $X$ . For each  $g \in G$ , let  $\text{Fix}(g) = \{x \in X : gx = x\}$  be the set of “fixed points” of  $g$ , and let  $N$  be the number of distinct orbits. Then

$$N = \frac{1}{|G|} \sum_{g \in G} |\text{Fix}(g)|.$$

So the number of orbits is the average number of fixed points of elements of  $G$ .

*Proof.* Define  $f : X \rightarrow \mathbb{R}$  by setting  $f(x) = 1/|Gx|$  for each  $x \in X$ . We will compute  $\sum_{x \in X} f(x)$  in two ways. Let  $\{O_1, \dots, O_N\}$  be the distinct orbits of the  $G$ -action. On one hand, grouping summands based on which orbit they are in, we get

$$\sum_{x \in X} f(x) = \sum_{i=1}^N \sum_{x \in O_i} f(x) = \sum_{i=1}^N \sum_{x \in O_i} \frac{1}{|O_i|} = \sum_{i=1}^N 1 = N.$$

On the other hand, 9.128 says that  $|Gx| = |G|/|\text{Stab}(x)|$ . Therefore

$$\begin{aligned} \sum_{x \in X} f(x) &= \sum_{x \in X} \frac{|\text{Stab}(x)|}{|G|} = \frac{1}{|G|} \sum_{x \in X} \sum_{g \in G} \chi(gx = x) \\ &= \frac{1}{|G|} \sum_{g \in G} \sum_{x \in X} \chi(gx = x) = \frac{1}{|G|} \sum_{g \in G} |\text{Fix}(g)|. \quad \square \end{aligned}$$

We are finally ready to solve the counting problems involving symmetry that were mentioned in the introduction to this chapter. The strategy is to introduce a set of objects  $X$  on which a certain group of symmetries acts. Each orbit of the group action consists of a set of objects in  $X$  that get identified with one another when symmetries are taken into account. So the solution to the counting problem is the number of orbits, which may be calculated by the formula of the previous theorem.

**9.141. Example: Counting Necklaces.** How many ways can we build a five-bead circular necklace if there are seven available types of gemstones (repeats allowed) and all rotations of a given necklace are considered equivalent? We can model the set of necklaces (before accounting for symmetries) by the set of words  $X = \{(y_1, y_2, y_3, y_4, y_5) : 1 \leq y_i \leq 7\}$ . Now let  $G = \langle (1, 2, 3, 4, 5) \rangle$  act on  $X$  by cyclically permuting positions (see 9.89). Every orbit of  $G$  consists of a set of necklaces that get identified with one another when symmetry is taken into account. To count the orbits, let us compute  $|\text{Fix}(g)|$  for each  $g \in G$ . First,  $\text{id} = (1)(2)(3)(4)(5)$  fixes every object in  $X$ , so  $|\text{Fix}(\text{id})| = |X| = 7^5$  by the product rule. Second, the generator  $g = (1, 2, 3, 4, 5)$  fixes  $(y_1, y_2, y_3, y_4, y_5)$  iff

$$(y_1, y_2, y_3, y_4, y_5) = (y_5, y_1, y_2, y_3, y_4).$$

Comparing coordinates, this holds iff  $y_1 = y_2 = y_3 = y_4 = y_5$  iff all the  $y_i$ 's are equal to one another. So  $|\text{Fix}((1, 2, 3, 4, 5))| = 7$  since there are seven choices for  $y_1$ . Next, what is  $|\text{Fix}(g^2)|$ ? We have  $g^2 = (1, 3, 5, 2, 4)$ , so that  $g^2$  fixes  $(y_1, y_2, y_3, y_4, y_5)$  iff

$$(y_1, y_2, y_3, y_4, y_5) = (y_4, y_5, y_1, y_2, y_3),$$

which holds iff  $y_1 = y_3 = y_5 = y_2 = y_4$ . So  $|\text{Fix}(g^2)| = 7$ . Similarly,  $|\text{Fix}(g^3)| = |\text{Fix}(g^4)| = 7$ , so the answer is

$$\frac{7^5 + 7 + 7 + 7 + 7}{5} = 3367.$$

Now suppose we are counting six-bead necklaces, identifying all rotations of a given necklace. Here, the group of symmetries is

$$G = \{\text{id}, (1, 2, 3, 4, 5, 6), (1, 3, 5)(2, 4, 6), (1, 4)(2, 5)(3, 6), (1, 5, 3)(2, 6, 4), (1, 6, 5, 4, 3, 2)\}.$$

As before,  $\text{id}$  has  $7^6$  fixed points, and each of the two six-cycles has 7 fixed points. What about  $\text{Fix}((1, 3, 5)(2, 4, 6))$ ? We have

$$(1, 3, 5)(2, 4, 6) * (y_1, y_2, y_3, y_4, y_5, y_6) = (y_5, y_6, y_1, y_2, y_3, y_4),$$

and this equals  $(y_1, \dots, y_6)$  iff  $y_1 = y_3 = y_5$  and  $y_2 = y_4 = y_6$ . Here there are 7 choices for  $y_1$ , 7 choices for  $y_2$ , and the remaining  $y_i$ 's are then forced. So  $|\text{Fix}((1, 3, 5)(2, 4, 6))| = 7^2$ . Likewise,  $|\text{Fix}((1, 5, 3)(2, 6, 4))| = 7^2$ . Similarly, we find that  $(y_1, \dots, y_6)$  is fixed by  $(1, 4)(2, 5)(3, 6)$  iff  $y_1 = y_4$  and  $y_2 = y_5$  and  $y_3 = y_6$ , so that there are  $7^3$  such fixed points. In each case,  $\text{Fix}(f)$  turned out to be  $7^{\text{cyc}(f)}$  where  $\text{cyc}(f)$  is the number of cycles in the complete cycle decomposition of  $f$  (including 1-cycles). The number of necklaces is

$$\frac{7^6 + 7 + 7^2 + 7^3 + 7^2 + 7}{6} = 19,684.$$

Now consider the question of counting five-bead necklaces using  $q$  types of beads, where rotations and reflections of a given necklace are considered equivalent. For this problem, the group of symmetries to use is the automorphism group of the cycle graph  $C_5$  (see 9.65). In addition to the five powers of  $(1, 2, 3, 4, 5)$ , this group contains the following five permutations corresponding to reflections of the necklace:

$$(1, 5)(2, 4)(3), (1, 4)(2, 3)(5), (1, 3)(4, 5)(2), (1, 2)(3, 5)(4), (2, 5)(3, 4)(1).$$

The reader may check that each of the five new permutations has  $q^3 = q^{\text{cyc}(f)}$  fixed points. For example, a necklace  $(y_1, \dots, y_5)$  is fixed by  $(1, 5)(2, 4)(3)$  iff  $y_1 = y_5$  ( $q$  choices) and  $y_2 = y_4$  ( $q$  choices) and  $y_3$  is arbitrary ( $q$  choices). So, the number of necklaces is

$$\frac{q^5 + 5q^3 + 4q^1}{10}.$$

The following general example can be used to solve many counting problems involving symmetry.

**9.142. Example: Counting Colorings under Symmetries.** Suppose  $V$  is a finite set of objects,  $C$  is a finite set of  $q$  colors, and  $G \subseteq \text{Sym}(V)$  is a group of symmetries of the objects  $V$ . (For example, if  $V$  is the vertex set of a graph, we could take  $G$  to be the automorphism group of the graph.)  $G$  acts on  $V$  via  $g \cdot x = g(x)$  for  $g \in G$  and  $x \in V$ . Now let  $X = {}^V C$  be the set of all functions  $f : V \rightarrow C$ . We think of a function  $f$  as a *coloring* of  $V$  such that  $x$  receives color  $f(x)$  for all  $x \in V$ . As we saw in 9.88,  $G$  acts on  $X$  via  $g * f = f \circ g^{-1}$  for  $g \in G$  and  $f \in X$ . Informally, if  $f$  assigns color  $c$  to object  $x$ , then  $g * f$  assigns color  $c$  to object  $g(x)$ . The  $G$ -orbits consist of colorings that get identified when we take into account the symmetries in  $G$ . So the number of colorings “up to symmetry” is  $\frac{1}{|G|} \sum_{g \in G} |\text{Fix}(g)|$ . In the previous example, we observed that  $|\text{Fix}(g)| = q^{\text{cyc}(g)}$ . To see why this holds in general, fix  $g \in G$  and write a complete cycle decomposition  $g = C_1 C_2 \cdots C_k$ , so  $k = \text{cyc}(g)$ . Let  $V_i$  be the elements appearing in cycle  $C_i$ , so  $V$  is the disjoint union of the sets  $V_i$ . Consider  $C_1$ , for example. Say  $C_1 = (x_1, x_2, \dots, x_s)$ , so that  $V_1 = \{x_1, \dots, x_s\}$ . Suppose  $f \in X$  is fixed by  $g$ , so  $f = g * f$ . Then

$$f(x_2) = (g * f)(x_2) = f(g^{-1}(x_2)) = f(x_1).$$



Similarly,  $f(x_3) = f(x_2)$ , and in general  $f(x_{j+1}) = f(x_j)$  for all  $j < s$ . It follows that  $f$  is constant on  $V_1$ . Similarly,  $f$  is constant on every  $V_i$  in the sense that  $f$  assigns the same color to every  $x \in V_i$ . This argument is reversible, so  $\text{Fix}(g)$  consists precisely of the colorings  $f \in {}^V C$  that are constant on each  $V_i$ . To build such an  $f$ , choose a common color for all the vertices in  $V_i$  (for  $1 \leq i \leq k$ ). By the product rule,  $|\text{Fix}(g)| = q^k = q^{\text{cyc}(g)}$  as claimed. Therefore, the answer to the counting problem is

$$\frac{1}{|G|} \sum_{g \in G} q^{\text{cyc}(g)}. \quad (9.8)$$

**9.143. Example: Counting Chessboards.** We now answer the question posed at the beginning of this chapter: how many ways can we color a  $5 \times 5$  chessboard with seven colors, if all rotations and reflections of a given colored board are considered the same? We apply the method of the preceding example. Let  $B = (V, E)$  be the graph that models the chessboard (Figure 9.3). Let  $C = \{1, 2, \dots, 7\}$  be the set of colors, and let  $X = {}^V C$  be the set of colorings before accounting for symmetry. The symmetry group  $G = \text{Aut}(B)$  was computed in 9.67. By inspecting the cycle decompositions for the eight elements  $g \in G$ , the answer follows from (9.8):

$$\frac{7^{25} + 7^7 + 7^{13} + 7^7 + 4 \cdot 7^{15}}{8} = 167,633,579,843,887,699,759.$$

## 9.19 Pólya's Formula

Consider the following variation of the chessboard coloring example: how many ways can we color a  $5 \times 5$  chessboard so that 10 squares are red, 12 are blue, and 3 are green, if all rotations and reflections of a colored board are equivalent? We can answer questions like this with the aid of “Pólya's Formula,” which extends Burnside's Lemma to *weighted* sets.

Let a finite group  $G$  act on a finite set  $X$ . Let  $\{O_1, \dots, O_N\}$  be the orbits of this action. Suppose each  $x \in X$  has a weight  $\text{wt}(x)$  in some polynomial ring  $R$ , and suppose that the weights are  $G$ -invariant:  $\text{wt}(g * x) = \text{wt}(x)$  for all  $g \in G$  and all  $x \in X$ . This condition implies that every object in a given  $G$ -orbit has the same weight. So we can assign a well-defined weight to each orbit by letting  $\text{wt}(O_i) = \text{wt}(x_i)$  for any  $x_i \in O_i$ . The next result lets us compute the generating function for the set of weighted orbits.

**9.144. Orbit-Counting Theorem for Weighted Sets.** With the above notation,

$$\sum_{i=1}^N \text{wt}(O_i) = \frac{1}{|G|} \sum_{g \in G} \sum_{x \in \text{Fix}(g)} \text{wt}(x).$$

So, the weighted sum of the orbits is the average over  $G$  of the weighted fixed point sets of elements of  $G$ .

*Proof.* We adapt the proof of the original orbit-counting theorem to include weights. Define  $f : X \rightarrow \mathbb{R}$  by setting  $f(x) = \text{wt}(x)/|Gx|$  for each  $x \in X$ . On one hand,

$$\sum_{x \in X} f(x) = \sum_{i=1}^N \sum_{x \in O_i} f(x) = \sum_{i=1}^N \sum_{x \in O_i} \frac{\text{wt}(x)}{|O_i|} = \sum_{i=1}^N \sum_{x \in O_i} \frac{\text{wt}(O_i)}{|O_i|} = \sum_{i=1}^N \text{wt}(O_i).$$

On the other hand, using  $|Gx| = |G|/|\text{Stab}(x)|$ , we get

$$\begin{aligned} \sum_{x \in X} f(x) &= \sum_{x \in X} \frac{|\text{Stab}(x)| \text{wt}(x)}{|G|} = \frac{1}{|G|} \sum_{x \in X} \sum_{g \in G} \chi(gx = x) \text{wt}(x) \\ &= \frac{1}{|G|} \sum_{g \in G} \sum_{x \in X} \chi(gx = x) \text{wt}(x) = \frac{1}{|G|} \sum_{g \in G} \sum_{x \in \text{Fix}(g)} \text{wt}(x). \quad \square \end{aligned}$$

We now extend the setup of 9.142 to count weighted colorings. We are given finite sets  $V$  and  $C = \{1, \dots, q\}$ , a subgroup  $G$  of  $\text{Sym}(V)$ , and the set of colorings  $X = {}^V C$ .  $G$  acts on  $X$  by permuting the domain:  $g * f = f \circ g^{-1}$  for  $g \in G$  and  $f \in X$ . We define a *weight* for a given coloring by setting

$$\text{wt}(f) = \prod_{x \in V} z_{f(x)} \in \mathbb{R}[z_1, z_2, \dots, z_q].$$

Note that  $\text{wt}(f) = z_1^{e_1} \cdots z_q^{e_q}$  iff  $f$  colors  $e_i$  of the objects in  $V$  with color  $i$ . We see that

$$\text{wt}(g * f) = \prod_{x \in V} z_{f(g^{-1}(x))} = \prod_{v \in V} z_{f(v)} = \text{wt}(f) \quad (g \in G, f \in X)$$

by making the change of variable  $v = g^{-1}(x)$ . So the weighted orbit-counting theorem is applicable. In the unweighted case (see 9.142), we found that  $|\text{Fix}(g)| = q^{\text{cyc}(g)}$  by arguing that  $f \in \text{Fix}(g)$  must be constant on each connected component  $V_1, \dots, V_k$  of the digraph of the permutation  $g$ . To take weights into account, let us construct such an  $f$  using the product rule for weighted sets. Suppose the components  $V_1, V_2, \dots, V_k$  have sizes  $n_1 \geq n_2 \geq \dots \geq n_k$  (so that  $\text{type}(g) = (n_1, n_2, \dots, n_k)$ ). First choose a common color for the  $n_1$  vertices in  $V_1$ . The generating function for this choice is  $z_1^{n_1} + z_2^{n_1} + \dots + z_q^{n_1}$ ; the term  $z_i^{n_1}$  arises by coloring all  $n_1$  vertices in  $V_1$  with color  $i$ . Second, choose a common color for the  $n_2$  vertices in  $V_2$ . The generating function for this choice is  $z_1^{n_2} + \dots + z_q^{n_2}$ . Continuing similarly, we arrive at the formula

$$\sum_{x \in \text{Fix}(g)} \text{wt}(x) = \prod_{i=1}^k (z_1^{n_i} + z_2^{n_i} + \dots + z_q^{n_i}).$$

We can abbreviate this formula by introducing the power-sum polynomials (which are studied in more detail in Chapter 10). For each integer  $k \geq 1$ , set  $p_k(z_1, \dots, z_q) = z_1^k + z_2^k + \dots + z_q^k$ . For each integer partition  $\mu = (\mu_1, \mu_2, \dots, \mu_k)$ , set  $p_\mu(z_1, \dots, z_q) = \prod_{i=1}^k p_{\mu_i}(z_1, \dots, z_q)$ . Then the weighted orbit-counting formula assumes the following form.

**9.145. Pólya's Formula.** With the above notation, the generating function for weighted colorings with  $q$  colors relative to the symmetry group  $G$  is

$$\sum_{i=1}^N \text{wt}(O_i) = \frac{1}{|G|} \sum_{g \in G} p_{\text{type}(g)}(z_1, z_2, \dots, z_q) \in \mathbb{R}[z_1, \dots, z_q].$$

The coefficient of  $z_1^{e_1} \cdots z_q^{e_q}$  in this polynomial is the number of colorings (taking the symmetries in  $G$  into account) in which color  $i$  is used  $e_i$  times.

**9.146. Example.** The generating function for five-bead necklaces using  $q$  types of beads (identifying all rotations and reflections of a given necklace) is

$$(p_{(1,1,1,1,1)} + 4p_{(5)} + 5p_{(2,2,1)})/10,$$

where all power-sum polynomials are evaluated at  $(z_1, \dots, z_q)$ .

**9.147. Example.** Let us use Pólya's formula to count  $5 \times 5$  chessboards with 10 red squares, 12 blue squares, and 3 green squares. We may as well take  $q = 3$  here. Consulting the cycle decompositions in 9.67 again, we find that the group  $G$  has one element of type  $(1^{25}) = (1, 1, \dots, 1)$ , two elements of type  $(4^6, 1)$ , one element of type  $(2^{12}, 1)$ , and four elements of type  $(2^{10}, 1^5)$ . Therefore,  $\sum_{i=1}^N \text{wt}(O_i)$  is given by

$$\frac{p_{(1^{25})}(z_1, z_2, z_3) + 2p_{(4^6, 1)}(z_1, z_2, z_3) + p_{(2^{12}, 1)}(z_1, z_2, z_3) + 4p_{(2^{10}, 1^5)}(z_1, z_2, z_3)}{8}.$$

Using a computer algebra system, we can compute this polynomial and extract the coefficient of  $z_1^{10}z_2^{12}z_3^3$ . The final answer is 185,937,878.

## Summary

Table 9.2 summarizes some definitions from group theory used in this chapter. Table 9.3 contains definitions pertinent to the theory of group actions.

- *Examples of Groups.* (i) additive commutative groups:  $\mathbb{Z}$ ,  $\mathbb{Q}$ ,  $\mathbb{R}$ ,  $\mathbb{C}$ , and  $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$  (under addition modulo  $n$ ); (ii) multiplicative commutative groups: invertible elements in  $\mathbb{Z}$ ,  $\mathbb{Q}$ ,  $\mathbb{R}$ ,  $\mathbb{C}$ , and  $\mathbb{Z}_n$ ; (iii) non-commutative groups: invertible matrices in  $M_n(R)$ , the group  $\text{Sym}(X)$  of bijections on  $X$  under composition, dihedral groups (automorphism groups of cycle graphs); (iv) constructions of groups: product groups (see 9.153), subgroups, quotient groups (see 9.205), cyclic subgroup generated by a group element, automorphism group of a graph, automorphism group of a group.
- *Basic Properties of Groups.* The identity of a group is unique, as is the inverse of each group element. In a group, there are left and right cancellation laws:  $(ax = ay) \Rightarrow (x = y)$  and  $(xa = ya) \Rightarrow (x = y)$ ; inverse rules:  $(x^{-1})^{-1} = x$  and  $(x_1 \cdots x_n)^{-1} = x_n^{-1} \cdots x_1^{-1}$ ; and the laws of exponents:  $x^{m+n} = x^m x^n$ ;  $(x^m)^n = x^{mn}$ ; and, when  $xy = yx$ ,  $(xy)^n = x^n y^n$ .
- *Notation for Permutations.* A bijection  $f \in S_n$  can be described in two-line form  $\begin{pmatrix} 1 & 2 & \cdots & n \\ f(1) & f(2) & \cdots & f(n) \end{pmatrix}$ , in one-line form  $[f(1), f(2), \dots, f(n)]$ , or in cycle notation. The cycle notation is obtained by listing the elements going around each directed cycle in the digraph of  $f$ , enclosing each cycle in parentheses, and optionally omitting cycles of length 1. The cycle notation for  $f$  is not unique.
- *Sorting, Inversions, and Sign.* A permutation  $w = w_1 w_2 \cdots w_n \in S_n$  can be sorted to the identity permutation  $\text{id} = 12 \cdots n$  by applying  $\text{inv}(w)$  basic transpositions to switch adjacent elements that are out of order. It follows that  $w$  can be written as the composition of  $\text{inv}(w)$  basic transpositions. Any factorization of  $w$  into a product of transpositions must involve an even number of terms when  $\text{sgn}(w) = +1$ , or an odd number when  $\text{sgn}(w) = -1$ . Sign is a group homomorphism:  $\text{sgn}(f \circ g) = \text{sgn}(f) \cdot \text{sgn}(g)$  for  $f, g \in S_n$ . The sign of a  $k$ -cycle is  $(-1)^{k-1}$ . For all  $f \in S_n$ ,  $\text{sgn}(f) = (-1)^{n - \text{cyc}(f)}$ .
- *Properties of Determinants.* The determinant of a matrix  $A \in M_n(R)$  is an  $R$ -multilinear, alternating function of the rows (resp. columns) of  $A$  such that  $\det(I_n) = 1_R$ . This means that  $\det(A)$  is an  $R$ -linear function of any given row when the other rows are fixed, and the determinant is zero if  $A$  has two equal rows; similarly for columns.

**TABLE 9.2**

Definitions in group theory.

Concept	Definition
group axioms for $(G, \star)$	$\begin{cases} \forall x, y \in G, x \star y \in G \text{ (closure)} \\ \forall x, y, z \in G, x \star (y \star z) = (x \star y) \star z \text{ (associativity)} \\ \exists e \in G, \forall x \in G, x \star e = x = e \star x \text{ (identity)} \\ \forall x \in G, \exists y \in G, x \star y = e = y \star x \text{ (inverses)} \end{cases}$
commutative group	group $G$ with $xy = yx$ for all $x, y \in G$
$H$ is a subgroup of $G$	$\begin{cases} e_G \in H \text{ (closure under identity)} \\ \forall a, b \in H, ab \in H \text{ (closure under operation)} \\ \forall a \in H, a^{-1} \in H \text{ (closure under inverses)} \end{cases}$
$H$ is <i>normal</i> in $G$ ( $H \trianglelefteq G$ )	$\forall g \in G, \forall h \in H, ghg^{-1} \in H$ (closure under conjugation)
exponent notation in $(G, \cdot)$	$x^0 = 1_G, x^{n+1} = x^n \cdot x, x^{-n} = (x^{-1})^n$ ( $n \geq 0$ )
multiple notation in $(G, +)$	$0x = 0_G, (n+1)x = nx + x, (-n)x = n(-x)$ ( $n \geq 0$ )
$k$ -cycle	$f \in \text{Sym}(X)$ of the form $(i_1, i_2, \dots, i_k)$ (cycle notation)
transposition	a 2-cycle $(i, j)$
basic transposition	a 2-cycle $(i, i+1)$ in $S_n$
$\text{cyc}(f)$	number of components in digraph of $f \in \text{Sym}(X)$
$\text{type}(f)$	list of cycle lengths of $f \in \text{Sym}(X)$ in decreasing order
$\text{inv}(w_1 \cdots w_n)$	number of $i < j$ with $w_i > w_j$
$\text{sgn}(w)$ for $w \in S_n$	$(-1)^{\text{inv}(w)}$
determinant of $A \in M_n(R)$	$\det(A) = \sum_{w \in S_n} \text{sgn}(w) \prod_{i=1}^n A(i, w(i))$
classical adjoint $\text{adj}(A)$	$\text{adj}(A)_{i,j} = (-1)^{i+j} \det(A[j i])$
cyclic subgroup $\langle x \rangle$	$\{x^n : n \in \mathbb{Z}\}$ or $\{nx : n \in \mathbb{Z}\}$ (additive notation)
cyclic group	group $G$ such that $G = \langle x \rangle$ for some $x \in G$
order of $x \in G$	least $n > 0$ with $x^n = e_G$ , or $\infty$ if no such $n$
graph automorphism of $K$	bijection on vertex set of $K$ preserving edges of $K$
group homomorphism	map $f : G \rightarrow H$ with $f(xy) = f(x)f(y)$ for all $x, y \in G$
kernel of hom. $f : G \rightarrow H$	$\ker(f) = \{x \in G : f(x) = e_H\}$
image of hom. $f : G \rightarrow H$	$\text{img}(f) = \{y \in H : y = f(x) \text{ for some } x \in G\}$
group isomorphism	bijective group homomorphism
group automorphism	group isomorphism from $G$ to itself
inner automorphism $C_g$	the automorphism $x \mapsto gxg^{-1}$ ( $g, x \in G$ )

We have  $\det(A^t) = \det(A)$ . For triangular or diagonal  $A$ ,  $\det(A) = \prod_{i=1}^n A(i, i)$ . The Laplace expansion for  $\det(A)$  along row  $k$  (resp. column  $k$ ) is

$$\det(A) = \sum_{j=1}^n (-1)^{j+k} A(k, j) \det(A[k|j]) = \sum_{i=1}^n (-1)^{i+k} A(i, k) \det(A[i|k]).$$

We have  $A(\text{adj } A) = (\det(A))I_n = (\text{adj } A)A$ , so that  $A^{-1} = (\det(A))^{-1} \text{adj}(A)$  when  $\det(A)$  is invertible in  $R$ . If  $m \leq n$ ,  $A$  is  $m \times n$ , and  $B$  is  $n \times m$ , the *Cauchy-Binet formula* says

$$\det(AB) = \sum_{1 \leq j_1 < j_2 < \cdots < j_m \leq n} \det(A^{j_1}, \dots, A^{j_m}) \det(B_{j_1}, \dots, B_{j_m}),$$

where  $A^j$  is the  $j$ th column of  $A$ , and  $B_j$  is the  $j$ th row of  $B$ . In particular,  $\det(AB) = \det(A) \det(B)$  for  $A, B \in M_n(R)$ .

- *Properties of Cyclic Groups.* Every cyclic group is commutative and isomorphic to  $\mathbb{Z}$

**TABLE 9.3**

Definitions in the theory of group actions.

Concept	Definition
action axioms for $G$ -set $X$	$\begin{cases} \forall g \in G, \forall x \in X, g * x \in X \text{ (closure)} \\ \forall x \in X, e_G * x = x \text{ (identity)} \\ \forall g, h \in G, \forall x \in X, g * (h * x) = (gh) * x \text{ (assoc.)} \end{cases}$
perm. representation of $G$ on $X$	group homomorphism $R : G \rightarrow \text{Sym}(X)$
$G$ -stable subset $Y$ of $X$	$\forall g \in G, \forall y \in Y, g * y \in Y$ (closure under action)
orbit of $x$ in $G$ -set $X$	$Gx = G * x = \{g * x : g \in G\}$
stabilizer of $x$ rel. to $G$ -set $X$	$\text{Stab}(x) = \{g \in G : g * x = x\} \leq G$
fixed points of $g$ in $G$ -set $X$	$\text{Fix}(g) = \{x \in X : g * x = x\}$
conjugacy class of $x$ in $G$	$\{gxg^{-1} : g \in G\}$
centralizer of $x$ in $G$	$C_G(x) = \{g \in G : gx = xg\} \leq G$
center of $G$	$Z(G) = \{g \in G : gx = xg \text{ for all } x \in G\} \trianglelefteq G$
normalizer of $H$ in $G$	$N_G(H) = \{g \in G : gHg^{-1} = H\} \leq G$
left coset of $H$	$xH = \{xh : h \in H\}$
right coset of $H$	$Hx = \{hx : h \in H\}$
set of left cosets $G/H$	for $H \leq G$ , $G/H = \{xH : x \in G\}$
index $[G : H]$	$[G : H] =  G/H $ = number of left cosets of $H$ in $G$

or  $\mathbb{Z}_n$  for some  $n \geq 1$ . More precisely, if  $G = \langle x \rangle$  is infinite, then  $f : \mathbb{Z} \rightarrow G$  given by  $f(i) = x^i$  for  $i \in \mathbb{Z}$  is a group isomorphism. If  $G = \langle x \rangle$  has size  $n$ , then  $g : \mathbb{Z}_n \rightarrow G$  given by  $g(i) = x^i$  for  $i \in \mathbb{Z}_n$  is a group isomorphism; moreover,  $x^m = e$  iff  $n$  divides  $m$ . Every subgroup of the additive group  $\mathbb{Z}$  has the form  $k\mathbb{Z}$  for a unique  $k \geq 0$ . Every subgroup of a cyclic group is cyclic.

- *Properties of Group Homomorphisms.* If  $f : G \rightarrow H$  is a group homomorphism, then  $\ker(f) \trianglelefteq G$  and  $\text{img}(f) \leq H$ . Moreover,  $f(x^n) = f(x)^n$  for all  $x \in G$  and  $n \in \mathbb{Z}$ . The composition of group homomorphisms (resp. isomorphisms) is a group homomorphism (resp. isomorphism), and the inverse of a group isomorphism is a group isomorphism.
- *Main Results on Group Actions.* Actions  $*$  of a group  $G$  on a set  $X$  correspond bijectively to permutation representations  $R : G \rightarrow \text{Sym}(X)$ , via the formula  $R(g)(x) = g * x$  for  $g \in G$  and  $x \in X$ . Every  $G$ -set  $X$  is the disjoint union of orbits; more precisely, each  $x \in X$  lies in a unique orbit  $Gx$ . The size of the orbit  $Gx$  is the index (number of cosets) of the stabilizer  $\text{Stab}(x)$  in  $G$ , which (for finite  $G$ ) is a divisor of  $|G|$ . The number of orbits is the average number of fixed points of elements of  $G$  (for  $G$  finite); this extends to weighted sets where the weight is constant on each orbit.
- *Examples of Group Actions.* A subgroup  $H$  of a group  $G$  acts on  $G$  by left multiplication ( $h * x = hx$ ), and by inverted right multiplication ( $h * x = xh^{-1}$ ), and by conjugation ( $h * x = h x h^{-1}$ ). The orbits of  $x$  under these respective actions are the right coset  $Hx$ , the left coset  $xH$ , and (when  $H = G$ ) the conjugacy class of  $x$  in  $G$ . Similarly,  $G$  (or its subgroups) act on the set of all subsets of  $G$  by left multiplication, and  $G$  acts by conjugation on the set of subgroups of  $G$ . The set of subsets of a fixed size  $k$  are also  $G$ -sets under these actions. Centralizers of elements and normalizers of subgroups are stabilizers under suitable actions, hence subgroups of  $G$ . Any subgroup  $G$  of  $\text{Sym}(X)$  acts on  $X$  by  $g * x = g(x)$  for  $g \in G$  and  $x \in X$ . For any set  $X$ ,  $S_n$  (or its subgroups) acts on  $X^n$  via  $f \cdot (x_1, \dots, x_n) = (x_{f^{-1}(1)}, \dots, x_{f^{-1}(n)})$ . For any subgroup  $H$  of  $G$ ,  $G$  acts on  $G/H$  via  $g * (xH) = (gx)H$  for  $g, x \in G$ .

- *Facts about Cosets.* Given a subgroup  $H$  of a group  $G$ ,  $G$  is the disjoint union of its left (resp. right) cosets, which all have the same cardinality as  $H$ . This implies Lagrange's theorem:  $|G| = |H| \cdot [G : H]$ , so that (for finite  $G$ ), the order and index of any subgroup of  $G$  are both divisors of  $|G|$ . To test equality of left cosets, one may check any of the following equivalent conditions:  $xH = yH$ ;  $x \in yH$ ;  $x = yh$  for some  $h \in H$ ;  $y^{-1}x \in H$ ;  $x^{-1}y \in H$ . Similarly,  $Hx = Hy$  iff  $xy^{-1} \in H$  iff  $yx^{-1} \in H$ . Left and right cosets coincide (i.e.,  $xH = Hx$  for all  $x \in G$ ) iff  $H$  is normal in  $G$  iff all conjugates  $xHx^{-1}$  equal  $H$  iff  $H$  is a union of conjugacy classes of  $G$ . Given a group homomorphism  $f : G \rightarrow L$  with kernel  $K$ ,  $Kx = xK = \{y \in G : f(y) = f(x)\}$  for all  $x \in G$ .
- *Conjugacy Classes.* Every group  $G$  is the disjoint union of its conjugacy classes, where the conjugacy class of  $x$  is  $\{gxg^{-1} : g \in G\}$ . Conjugacy classes need not all have the same size. The size of the conjugacy class of  $x$  is the index  $[G : C_G(x)]$ , where  $C_G(x)$  is the subgroup  $\{y \in G : xy = yx\}$ ; this index is a divisor of  $|G|$  for  $G$  finite. For  $x \in G$ , the conjugacy class of  $x$  has size 1 iff  $x$  is in the center  $Z(G)$ . This can be used to show that groups  $G$  of size  $p^n$  (where  $p$  is prime and  $n \geq 1$ ) have  $|Z(G)| > 1$ . Each conjugacy class of  $S_n$  consists of those  $f \in S_n$  with a given cycle type  $\mu \in \text{Par}(n)$ . This follows from the fact that the cycle notation for  $gfg^{-1}$  is the cycle notation for  $f$  with each value  $x$  replaced by  $g(x)$ . The size of the conjugacy class indexed by  $\mu$  is  $n!/z_\mu$ .
- *Cayley's Theorem on Permutation Representations.* Every group  $G$  is isomorphic to a subgroup of  $\text{Sym}(G)$ , via the homomorphism sending  $g \in G$  to the left multiplication  $L_g = (x \mapsto gx : x \in G)$ . Every  $n$ -element group is isomorphic to a subgroup of  $S_n$ .
- *Theorems Provable by Group Actions.* (i) Fermat's Little Theorem:  $a^p \equiv a \pmod{p}$  for  $a \in \mathbb{N}^+$  and  $p$  prime. (ii) Cauchy's Theorem: If  $G$  is a group and  $p$  is a prime divisor of  $|G|$ , then there exists  $x \in G$  of order  $p$ . (iii) Lucas' Congruence: For  $0 \leq k \leq n$  and prime  $p$ ,  $\binom{n}{k} \equiv \prod_{i \geq 0} \binom{n_i}{k_i} \pmod{p}$ , where the  $n_i$  and  $k_i$  are the base- $p$  digits of  $n$  and  $k$ . (iv) Sylow's First Theorem: If  $G$  is a group and  $|G|$  has prime factorization  $|G| = p_1^{n_1} \cdots p_k^{n_k}$ , then  $G$  has a subgroup of size  $p_i^{n_i}$  for  $1 \leq i \leq k$ .
- *Counting Colorings under Symmetries.* Given a finite set  $V$ , a group of symmetries  $G \leq \text{Sym}(V)$ , and a set  $C$  of  $q$  colors, the number of colorings  $f : V \rightarrow C$  taking symmetries into account is  $|G|^{-1} \sum_{g \in G} q^{\text{cyc}(g)}$ . If the colors are weighted using  $z_1, \dots, z_q$ , the generating function for weighted colorings is given by Pólya's formula

$$\frac{1}{|G|} \sum_{g \in G} p_{\text{type}(g)}(z_1, \dots, z_q),$$

where  $p_\mu$  is a power-sum symmetric polynomial. The coefficient of  $z_1^{e_1} \cdots z_q^{e_q}$  gives the number of colorings (taking the symmetries in  $G$  into account) where color  $i$  is used  $e_i$  times.

## Exercises

**9.148.** Let  $X$  be a set with more than one element. Define  $a \star b = b$  for all  $a, b \in X$ . (a) Prove that  $(X, \star)$  satisfies the closure axiom and associativity axiom in 9.1. (b) Does there exist  $e \in X$  such that  $e \star x = x$  for all  $x \in X$ ? If so, is this  $e$  unique? (c) Does there exist  $e \in X$  such that  $x \star e = x$  for all  $x \in X$ ? If so, is this  $e$  unique? (d) Is  $(X, \star)$  a group?

**9.149.** Let  $G$  be the set of odd integers. For all  $x, y \in G$ , define  $x \star y = x + y + 5$ . Prove that  $(G, \star)$  is a commutative group.

**9.150.** Let  $G$  be the set of real numbers unequal to 1. For each  $a, b \in G$ , define  $a \star b = a + b - ab$ . Prove that  $(G, \star)$  is a commutative group.

**9.151.** Assume  $(G, \star)$  is a group such that  $x \star x = e$  for all  $x \in G$ , where  $e$  is the identity element of  $G$ . Prove that  $G$  is commutative.

**9.152.** Let  $(G, \star)$  be a group. Define  $\bullet : G \times G \rightarrow G$  by setting  $a \bullet b = b \star a$  for all  $a, b \in G$ . Prove that  $(G, \bullet)$  is a group.

**9.153. Product Groups.** Let  $(G, \star)$  and  $(H, \bullet)$  be groups. (a) Show that  $G \times H$  becomes a group under the operation  $(g_1, h_1) * (g_2, h_2) = (g_1 \star g_2, h_1 \bullet h_2)$  for  $g_1, g_2 \in G, h_1, h_2 \in H$ . (b) Show  $G \times H$  is commutative iff  $G$  and  $H$  are commutative.

**9.154.** Prove the associative axiom for  $(\mathbb{Z}_n, \oplus)$  by verifying (9.1).

**9.155.** Suppose  $G$  is a set,  $\star : G \times G \rightarrow G$  is associative, and there exists  $e \in G$  such that for all  $x \in G$ ,  $e \star x = x$  and there is  $y \in G$  with  $y \star x = e$ . Prove  $(G, \star)$  is a group.

**9.156.** For  $x, y$  in a group  $G$ , define the *commutator*  $[x, y] = xyx^{-1}y^{-1}$ , and let  $C_x(y) = xyx^{-1}$  (conjugation by  $x$ ). Verify that the following identities hold for all  $x, y, z \in G$ : (a)  $[x, y]^{-1} = [y, x]$ ; (b)  $[x, yz] = [x, y]C_y([x, z])$ ; (c)  $[x, yz][y, zx][z, xy] = e_G$ ; (d)  $[[x, y], C_y(z)][[y, z], C_z(x)][[z, x], C_x(y)] = e_G$ .

**9.157.** Give complete proofs of the three laws of exponents in 9.10.

**9.158.** Let  $G$  be a group. For each  $g \in G$ , define a function  $R_g : G \rightarrow G$  by setting  $R_g(x) = xg$  for each  $x \in G$ .  $R_g$  is called “right multiplication by  $g$ .” (a) Prove that  $R_g$  is one-to-one and onto. (b) Prove that  $R_e = \text{id}_G$  (where  $e$  is the identity of  $G$ ) and  $R_g \circ R_h = R_{hg}$  for all  $g, h \in G$ . (c) Point out why  $R_g$  is an element of  $\text{Sym}(G)$ . Give two answers, one based on (a) and one based on (b). (d) Define  $\phi : G \rightarrow \text{Sym}(G)$  by setting  $\phi(g) = R_g$  for  $g \in G$ . Prove that  $\phi$  is one-to-one. (e) Prove that for all  $g, h \in G$ ,  $L_g \circ R_h = R_h \circ L_g$  (where  $L_g$  is left multiplication by  $g$ ).

**9.159.** Let  $G$  be a group. (a) Prove that for all  $a, b \in G$ , there exists a unique  $x \in G$  with  $ax = b$ . (b) Prove that in the multiplication table for a group  $G$ , every group element appears exactly once in each row and column.

**9.160.** A certain group  $(G, \star)$  has a multiplication table that has been partly filled in below:

$\star$	1	2	3	4
1	4			
2		1		
3			1	
4				

Use properties of groups to fill in the rest of the table.

**9.161.** Let  $f, g \in S_8$  be given in one-line form by  $f = [3, 2, 7, 5, 1, 4, 8, 6]$  and  $g = [4, 5, 1, 3, 2, 6, 8, 7]$ . (a) Write  $f$  and  $g$  in cycle notation. (b) Compute  $f \circ g$ ,  $g \circ f$ ,  $g \circ g$ , and  $f^{-1}$ , giving final answers in one-line form.

**9.162.** Let  $h = [4, 1, 3, 6, 5, 2]$  in one-line form. Compute  $\text{inv}(h)$  and  $\text{sgn}(h)$ . Write  $h$  as a product of  $\text{inv}(h)$  basic transpositions.

**9.163.** Let  $f = (1, 3, 6)(2, 8)(4)(5, 7)$  and  $g = (5, 4, 3, 2, 1)(7, 8)$ . (a) Compute  $fg$ ,  $gf$ ,  $fgf^{-1}$ , and  $gfg^{-1}$ , giving all answers in cycle notation. (b) Compute  $\text{sgn}(f)$  and  $\text{sgn}(g)$  without counting inversions. (c) Find an  $h \in S_8$  such that  $hfh^{-1} = (1, 2, 3)(4, 5)(6)(7, 8)$ ; give the answer in two-line form.

**9.164.** Suppose that  $f \in S_n$  has cycle type  $\mu = (\mu_1, \dots, \mu_k)$ . What is the order of  $f$ ?

**9.165.** The *support* of a bijection  $f \in \text{Sym}(X)$  is the set  $\text{supp}(f) = \{x \in X : f(x) \neq x\}$ . Two permutations  $f, g \in \text{Sym}(X)$  are called *disjoint* iff  $\text{supp}(f) \cap \text{supp}(g) = \emptyset$ . (a) Prove that for all  $x \in X$  and  $f \in \text{Sym}(X)$ ,  $x \in \text{supp}(f)$  implies  $f(x) \in \text{supp}(f)$ . (b) Prove that *disjoint permutations commute*, i.e., for all disjoint  $f, g \in \text{Sym}(X)$ ,  $f \circ g = g \circ f$ . (c) Suppose  $f \in \text{Sym}(X)$  is given in cycle notation by  $f = C_1 C_2 \cdots C_k$ , where the  $C_i$  are cycles involving pairwise disjoint subsets of  $X$ . Show that the  $C_i$ 's commute with one another, and prove carefully that  $f = C_1 \circ C_2 \circ \cdots \circ C_k$  (cf. 9.19).

**9.166.** Prove 9.27.

**9.167.** (a) Verify the formula  $(i_1, i_2, \dots, i_k) = (i_1, i_2) \circ (i_2, i_3) \circ (i_3, i_4) \circ \cdots \circ (i_{k-1}, i_k)$  used in the proof of 9.33. (b) Prove that every transposition has sign  $-1$  by finding an explicit formula for  $(i, j)$  as a product of an odd number of basic transpositions (which have sign  $-1$  by 9.26 with  $w = \text{id}$ ).

**9.168.** Given  $f \in S_n$ , what is the relationship between the one-line forms of  $f$  and  $f \circ (i, j)$ ? What about  $f$  and  $(i, j) \circ f$ ?

**9.169.** Let  $f \in S_n$  and  $h = (i, i+1) \circ f$ . (a) Prove an analogue of 9.26 relating  $\text{inv}(f)$  to  $\text{inv}(h)$  and  $\text{sgn}(f)$  to  $\text{sgn}(h)$ . (b) Use (a) to give another proof of the formula  $\text{sgn}(f \circ g) = \text{sgn}(f)\text{sgn}(g)$  that proceeds by induction on  $\text{inv}(f)$ .

**9.170.** Prove that for  $n \geq 3$ , every  $f \in A_n$  can be written as a product of 3-cycles.

**9.171.** Suppose an  $n \times n$  matrix  $A$  is given in block form as  $A = \begin{bmatrix} B & 0 \\ C & D \end{bmatrix}$ , where  $B$  is  $k \times k$ ,  $C$  is  $(n-k) \times k$ ,  $D$  is  $(n-k) \times (n-k)$ , and  $0$  denotes a  $k \times (n-k)$  block of zeroes. Prove that  $\det(A) = \det(B)\det(D)$ .

**9.172. Algorithmic Complexity of Determinant Evaluation.** Let  $A \in M_n(F)$  where  $F$  is a field. (a) How many additions and multiplications in  $F$  are needed to compute  $\det(A)$  directly from 9.37? (b) How many additions and multiplications in  $F$  are needed to compute  $\det(A)$  recursively, using 9.48? (c) Explain how to use 9.41 and 9.47 to compute  $\det(A)$  efficiently (using about  $cn^3$  field operations for some constant  $c$ ).

**9.173. Permanents.** The *permanent* of an  $n \times n$  matrix  $A \in M_n(R)$  is defined as  $\text{per}(A) = \sum_{w \in S_n} \prod_{i=1}^n A(i, w(i))$ . Prove the following facts about permanents: (a)  $\text{per}(A^t) = \text{per}(A)$ ; (b) if  $A$  is diagonal, then  $\text{per}(A) = \prod_{i=1}^n A(i, i)$ ; (c)  $\text{per}(I_n) = 1_R$ ; (d)  $\text{per}(A)$  is an  $R$ -multilinear function of the rows and columns of  $A$  (cf. 9.45); (e) if  $B$  is obtained from  $A$  by permuting the rows in any fashion, then  $\text{per}(B) = \text{per}(A)$ .

**9.174.** State and prove analogues for permanents of the Laplace expansions in 9.48.

**9.175.** Verify the characterization of  $R$ -linear maps stated in 9.44.

**9.176.** Complete the proof of 9.50 by showing that  $(\text{adj } A)A = \det(A)I_n$ .

**9.177. Cramer's Rule.** Let  $A \in M_n(R)$  where  $\det(A)$  is invertible in  $R$ , let  $b$  be a given  $n \times 1$  vector, and let  $x = [x_1 \cdots x_n]^t$ . Show that the unique solution of the linear system  $Ax = b$  is given by  $x_i = \det(A_i)/\det(A)$ , where  $A_i$  is the matrix obtained from  $A$  by replacing the  $i$ th column by  $b$ .



**9.178.** Verify the Cauchy-Binet formula for the matrices

$$A = \begin{bmatrix} 2 & 1 & 0 & 3 \\ 1 & -1 & 1 & 2 \\ 4 & 0 & 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 1 & -1 \\ 0 & 2 & 5 \\ 1 & 1 & 4 \\ -2 & 0 & -1 \end{bmatrix}.$$

**9.179.** Consider a function  $w : \{1, 2, \dots, k\} \rightarrow \{1, 2, \dots, n\}$ , which we regard as a word  $w = w_1 w_2 \cdots w_k$ . Show that there exist basic transpositions  $t_1, \dots, t_m \in S_k$  such that  $w \circ (t_1 t_2 \cdots t_m)$  is a weakly increasing word, and the minimum possible value of  $m$  is  $\text{inv}(w) = \sum_{i < j} \chi(w_i > w_j)$ .

**9.180.** Let  $A$  and  $B$  be  $n \times n$  matrices. Prove that  $\det(AB) = \det(A) \det(B)$  by imitating (and simplifying) the proof of the Cauchy-Binet formula 9.53.

**9.181.** Verify all the assertions in 9.57.

**9.182.** Let  $x$  be an element of a group  $G$ , written multiplicatively. Use the laws of exponents to verify that  $\langle x \rangle = \{x^n : n \in \mathbb{Z}\}$  is a subgroup of  $G$ , as stated in 9.60.

**9.183. Subgroup Generated by a Set.** Let  $S$  be a nonempty subset of a group  $G$ . Let  $\langle S \rangle$  be the set of elements of  $G$  of the form  $x_1 x_2 \cdots x_n$ , where  $n \in \mathbb{N}^+$  and, for  $1 \leq i \leq n$ , either  $x_i \in S$  or  $x_i^{-1} \in S$ . Prove that  $\langle S \rangle \leq G$ , and for all  $T$  with  $S \subseteq T \leq G$ ,  $\langle S \rangle \leq T$ .

**9.184.** Prove that every subgroup of a cyclic group is cyclic.

**9.185.** For subsets  $S$  and  $T$  of a multiplicative group  $G$ , define  $ST = \{st : s \in S, t \in T\}$ . (a) Show that if  $S \trianglelefteq G$  and  $T \leq G$ , then  $ST = TS$  and  $ST \leq G$ . Give an example to show  $ST$  may not be normal in  $G$ . (b) Show that if  $S \trianglelefteq G$  and  $T \trianglelefteq G$ , then  $ST \trianglelefteq G$ . (c) Give an example of a group  $G$  and subgroups  $S$  and  $T$  such that  $ST$  is not a subgroup of  $G$ .

**9.186.** Let  $S$  and  $T$  be finite subgroups of a group  $G$ . Prove that  $|S| \cdot |T| = |ST| \cdot |S \cap T|$ .

**9.187.** Assume that  $G$  is a group and  $H \leq G$ . Let  $H^{-1} = \{h^{-1} : h \in H\}$ . (a) Show that  $HH = H^{-1} = HH^{-1} = H$ . (b) Prove that  $H \trianglelefteq G$  iff  $gHg^{-1} = H$  for all  $g \in G$ .

**9.188.** Show that a subgroup  $H$  of a group  $G$  is normal in  $G$  iff  $H$  is a union of conjugacy classes of  $G$ .

**9.189.** Find all the subgroups of  $S_4$ . Which subgroups are normal? Confirm that Sylow's theorem 9.139 is true for this group.

**9.190.** Find all *normal* subgroups of  $S_5$ , and prove that you have found them all (Lagrange's theorem and 9.188 can be helpful here).

**9.191.** Suppose  $H$  is a *finite*, nonempty subset of a group  $G$  such that  $xy \in H$  for all  $x, y \in H$ . Prove that  $H \leq G$ . Give an example to show this result may not be true if  $H$  is not finite.

**9.192.** Given any simple graph or digraph  $K$  with vertex set  $X$ , show that  $\text{Aut}(K)$  is a subgroup of  $\text{Sym}(X)$ .

**9.193.** Determine the automorphism groups of the following graphs and digraphs: (a) the path graph  $P_n$  (see 3.124); (b) the complete graph  $K_n$  (see 3.124); (c) the empty graph on  $\{1, 2, \dots, n\}$  with no edges; (d) the directed cycle with vertex set  $\{1, 2, \dots, n\}$  and edges  $(n, 1)$  and  $(i, i + 1)$  for  $i < n$ ; (e) the simple graph with vertex set  $\{\pm 1, \pm 2, \dots, \pm n\}$  and edge set  $\{\{i, -i\} : 1 \leq i \leq n\}$ .

**9.194.** Let  $K$  be the Petersen graph (defined in 3.215). (a) Given two paths  $P = (y_0, y_1, y_2, y_3)$  and  $Q = (z_0, z_1, z_2, z_3)$  in  $K$ , prove that there exists a unique automorphism of  $K$  that maps  $y_i$  to  $z_i$  for  $0 \leq i \leq 3$ . (b) Prove that  $K$  has exactly  $5! = 120$  automorphisms. (c) Is  $\text{Aut}(K)$  isomorphic to  $S_5$ ?

**9.195.** Let  $Q_k$  be the simple graph with vertex set  $V = \{0, 1\}^k$  and edge set  $E = \{(v, w) \in V : v, w \text{ differ in exactly one position}\}$ .  $Q_k$  is called a  $k$ -dimensional hypercube. (a) Compute  $|V(Q_k)|$ ,  $|E(Q_k)|$ , and  $\deg(Q_k)$ . (b) Show that  $Q_k$  has exactly  $\binom{k}{i} 2^{k-i}$  induced subgraphs isomorphic to  $Q_i$ . (c) Find all the automorphisms of  $Q_k$ . How many are there?

**9.196.** (a) Construct an *undirected* graph whose automorphism group has size three. What is the minimum number of vertices in such a graph? (b) For each  $n \geq 1$ , construct an undirected graph whose automorphism group is cyclic of size  $n$ .

**9.197.** Let  $G$  be a simple graph with connected components  $C_1, \dots, C_k$ . What is the relation between  $|\text{Aut}(G)|$  and  $(|\text{Aut}(C_i)| : 1 \leq i \leq k)$ ?

**9.198.** Let  $f : G \rightarrow H$  be a group homomorphism. (a) Show that if  $K \leq G$ , then  $f[K] = \{f(x) : x \in K\}$  is a subgroup of  $H$ . If  $K \trianglelefteq G$ , must  $f[K]$  be normal in  $H$ ? (b) Show that if  $L \leq H$ , then  $f^{-1}[L] = \{x \in G : f(x) \in L\}$  is a subgroup of  $G$ . If  $L \trianglelefteq H$ , must  $f^{-1}[L]$  be normal in  $G$ ? (c) Deduce from (a) and (b) that the kernel and image of a group homomorphism are subgroups.

**9.199.** Show that the group of nonzero complex numbers under multiplication is isomorphic to the product of the subgroups  $\mathbb{R}^+$  and  $\{z \in \mathbb{C} : |z| = 1\}$ .

**9.200.** Give examples of four non-isomorphic groups of size 12.

**9.201.** Suppose  $G$  is a commutative group with subgroups  $H$  and  $K$ , such that  $G = HK$  and  $H \cap K = \{e_G\}$ . (a) Prove that the map  $(h, k) \mapsto hk$  is a group isomorphism from  $H \times K$  onto  $G$ . (b) Does any analogous result hold if  $G$  is not commutative? What if  $H$  and  $K$  are normal in  $G$ ?

**9.202.** (a) Let  $G$  be a group and  $x \in G$ . Show there exists a unique group homomorphism  $f : \mathbb{Z} \rightarrow G$  with  $f(1) = x$ . (b) Use (a) to determine the group  $\text{Aut}(\mathbb{Z})$ .

**9.203.** (a) Suppose  $G$  is a group,  $x \in G$ , and  $x^n = e_G$  for some  $n \geq 2$ . Show there exists a unique group homomorphism  $f : \mathbb{Z}_n \rightarrow G$  with  $f(1) = x$ . (b) Use (a) to prove that  $\text{Aut}(\mathbb{Z}_n)$  is isomorphic to the group  $\mathbb{Z}_n^*$  of invertible elements of  $\mathbb{Z}_n$  under multiplication modulo  $n$ .

**9.204. Properties of Order.** Let  $G$  be a group and  $x \in G$ . (a) Prove  $x$  and  $x^{-1}$  have the same order. (b) Show that if  $x$  has infinite order, then so does  $x^i$  for all nonzero integers  $i$ . (c) Suppose  $x$  has finite order  $n$ . Show that the order of  $x^k$  is  $n/\gcd(k, n)$  for all  $k \in \mathbb{Z}$ . (d) Show that if  $f : G \rightarrow H$  is a group isomorphism, then  $x$  and  $f(x)$  have the same order. What can be said if  $f$  is only a group homomorphism?

**9.205. Quotient Groups.** (a) Suppose  $H$  is a *normal* subgroup of  $G$ . Show that the set  $G/H$  of left cosets of  $H$  in  $G$  becomes a group of size  $[G : H]$  if we define  $(xH) \star (yH) = (xy)H$  for all  $x, y \in G$ . (One must first show that this operation is *well-defined*: i.e., for all  $x_1, x_2, y_1, y_2 \in G$ ,  $x_1H = x_2H$  and  $y_1H = y_2H$  imply  $x_1y_1H = x_2y_2H$ . For this, use the coset equality theorem.) (b) With the notation in (a), define  $\pi : G \rightarrow G/H$  by  $\pi(x) = xH$  for  $x \in G$ . Show that  $\pi$  is a surjective group homomorphism with kernel  $H$ . (c) Let  $H = \{\text{id}, (1, 2)\} \leq S_3$ . Find  $x_1, x_2, y_1, y_2 \in S_3$  with  $x_1H = x_2H$  and  $y_1H = y_2H$ , but  $x_1y_1H \neq x_2y_2H$ . This shows that normality of  $H$  is needed for the product in (a) to be well defined.

**9.206.** Let  $H$  be a normal subgroup of a group  $G$ . (a) Prove that  $G/H$  is commutative if  $G$  is commutative. (b) Prove that  $G/H$  is cyclic if  $G$  is cyclic. (c) Does the converse of (a) or (b) hold? Explain.

**9.207. Fundamental Homomorphism Theorem for Groups.** Suppose  $G$  and  $H$  are groups and  $f : G \rightarrow H$  is a group homomorphism. Let  $K = \{x \in G : f(x) = e_H\}$  be the kernel of  $f$ , and let  $I = \{y \in H : \exists x \in G, y = f(x)\}$  be the image of  $f$ . Show that  $K \trianglelefteq G$ ,  $I \leq H$  and there exists a unique group isomorphism  $\bar{f} : G/K \rightarrow I$  given by  $\bar{f}(xK) = f(x)$  for  $x \in G$ .

**9.208. Universal Mapping Property for Quotient Groups.** Let  $G$  be a group with normal subgroup  $N$ , let  $\pi : G \rightarrow G/N$  be the homomorphism  $\pi(x) = xN$  for  $x \in G$ , and let  $H$  be any group. (a) Show that if  $h : G/N \rightarrow H$  is a group homomorphism, then  $h \circ \pi$  is a group homomorphism from  $G$  to  $H$  sending each  $n \in N$  to  $e_H$ . (b) Conversely, given any group homomorphism  $f : G \rightarrow H$  such that  $f(n) = e_H$  for all  $n \in N$ , show that there exists a unique group homomorphism  $h : G/N \rightarrow H$  such that  $f = h \circ \pi$ . (c) Conclude that the map  $h \mapsto h \circ \pi$  is a *bijection* from the set of all group homomorphisms from  $G/N$  to  $H$  to the set of all group homomorphisms from  $G$  to  $H$  that map everything in  $N$  to  $e_H$ .

**9.209. Diamond Isomorphism Theorem for Groups.** Suppose  $G$  is a group,  $S \trianglelefteq G$ , and  $T \leq G$ . Show (cf. 9.185)  $TS = ST \leq G$ ,  $S \trianglelefteq TS$ ,  $(S \cap T) \trianglelefteq T$ , and there is a well-defined group isomorphism  $f : T/(S \cap T) \rightarrow (TS)/S$  given by  $f(x(S \cap T)) = xS$  for all  $x \in T$ . Use this to give another solution of 9.186 in the case where  $S$  is *normal* in  $G$ .

**9.210. Double-Quotient Isomorphism Theorem for Groups.** Assume  $A \leq B \leq C$  are groups with  $A$  and  $B$  both normal in  $C$ . Show that  $A \trianglelefteq B$ ,  $B/A \trianglelefteq C/A$ , and  $(C/A)/(B/A)$  is isomorphic to  $C/B$  via the map  $(xA)B/A \mapsto xB$  for  $x \in C$ .

**9.211. Correspondence Theorem for Quotient Groups.** Let  $H$  be a normal subgroup of a group  $G$ . Let  $X$  be the set of subgroups of  $G$  containing  $H$ , and let  $Y$  be the set of subgroups of  $G/H$ . Show that the map  $L \mapsto L/H = \{xH : x \in L\}$  is an inclusion-preserving bijection of  $X$  onto  $Y$  with inverse  $M \mapsto \{x \in G : xH \in M\}$ . If  $L$  maps to  $M$  under this correspondence, show that  $[G : L] = [G/H : M]$ , that  $[L : H] = |M|$ , that  $L \trianglelefteq G$  iff  $M \trianglelefteq G/H$ , and that  $G/L$  is isomorphic to  $(G/H)/M$  whenever  $L \trianglelefteq G$ .

**9.212.** Let  $G$  be a non-commutative group. Show that the rule  $g \cdot x = xg$  (for  $g, x \in G$ ) does not define a left action of  $G$  on the set  $G$ .

**9.213.** Let  $G$  act on itself by conjugation:  $g * x = gxg^{-1}$  for  $g, x \in G$ . Verify that the axioms for a left group action are satisfied.

**9.214.** Let  $(X, *)$  be a  $G$ -set and  $f : K \rightarrow G$  a group homomorphism. Verify the  $K$ -set axioms for the action  $k \bullet x = f(k) * x$  ( $k \in K$ ,  $x \in X$ ).

**9.215.** Suppose  $* : G \times X \rightarrow X$  is a group action. (a) Show that  $\mathcal{P}(X)$  is a  $G$ -set via the action  $g \bullet S = \{g * s : s \in S\}$  for  $g \in G$  and  $S \in \mathcal{P}(X)$ . (b) For fixed  $k$ , show that the set of all  $k$ -element subsets of  $X$  is a  $G$ -stable subset of  $\mathcal{P}(X)$ .

**9.216.** Verify the action axioms for the action of  $S_n$  on  $V$  in 9.87.

**9.217.** Suppose  $(X, *)$  is a  $G$ -set and  $W$  is a set. Show that the set of functions  $F : W \rightarrow X$  is a  $G$ -set via the action  $(g \bullet F)(w) = g * (F(w))$  for all  $g \in G$ ,  $F \in {}^W X$ , and  $w \in W$ .

**9.218.** Let a subgroup  $H$  of a group  $G$  act on  $G$  via  $h * x = xh^{-1}$  for  $h \in H$  and  $x \in G$ . Show that the orbit  $H * x$  is the left coset  $xH$ , for  $x \in G$ .

**9.219.** (a) Suppose  $f : X \rightarrow Y$  is a bijection. Show that the map  $T : \text{Sym}(X) \rightarrow \text{Sym}(Y)$  given by  $T(g) = f \circ g \circ f^{-1}$  for  $g \in \text{Sym}(X)$  is a group isomorphism. (b) Use (a) and Cayley's theorem to conclude that every  $n$ -element group is isomorphic to a subgroup of  $S_n$ .

**9.220.** Let a group  $G$  act on a set  $X$ . Show that  $\bigcap_{x \in X} \text{Stab}(x)$  is a *normal* subgroup of  $G$ . Give an example to show that a stabilizer subgroup  $\text{Stab}(x)$  may not be normal in  $G$ .

**9.221.** Let  $G$  act on itself by conjugation. (a) By considering the associated permutation representation and using the fundamental homomorphism theorem 9.207, deduce that  $G/Z(G)$  is isomorphic to the subgroup of inner automorphisms in  $\text{Aut}(G)$ . (b) Show that the subgroup of inner automorphisms is *normal* in  $\text{Aut}(G)$ .

**9.222.** Let  $(\mathbb{R}, +)$  act on  $\mathbb{R}^2$  (viewed as column vectors) by the rule

$$\theta * \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (\theta, x, y \in \mathbb{R}).$$

Verify that this is an action, and describe the orbit and stabilizer of each point in  $\mathbb{R}^2$ .

**9.223.** Let  $f \in S_n$ , and let  $\langle f \rangle$  act on  $\{1, 2, \dots, n\}$  via  $g \cdot x = g(x)$  for all  $g \in \langle f \rangle$  and  $x \in \{1, 2, \dots, n\}$ . Prove that the orbits of this action are the connected components of the digraph of  $f$ .

**9.224.** Suppose  $X$  is a  $G$ -set and  $x, y \in X$ . Without appealing to equivalence relations, give a direct proof that  $Gx \cap Gy \neq \emptyset$  implies  $Gx = Gy$ .

**9.225.** Let  $*$  be a right action of a group  $G$  on a set  $X$ . (a) Prove that  $X$  is the disjoint union of orbits  $x * G$ . (b) Prove that  $|x * G| = [G : \text{Stab}(x)]$ , where  $\text{Stab}(x) = \{g \in G : x * g = x\}$ .

**9.226.** State and prove a version of the coset equality theorem 9.111 for right cosets.

**9.227.** Let  $G$  be a group with subgroup  $H$ . Prove that the map  $T(xH) = Hx^{-1}$  for  $x \in G$  is a well-defined bijection from the set of left cosets of  $H$  in  $G$  onto the set of right cosets of  $H$  in  $G$ .

**9.228.** Let  $X$  be a  $G$ -set. For  $x \in X$  and  $g \in G$ , prove that  $g\text{Stab}(x) = \{h \in G : h * x = g * x\}$ . (This shows that each left coset of the stabilizer of  $x$  consists of those group elements sending  $x$  to a particular element in its orbit  $Gx$ . Compare to 9.120.)

**9.229.** Let  $G$  be a group with subgroup  $H$ . Prove the following facts about the normalizer of  $H$  in  $G$  (see 9.126). (a)  $N_G(H)$  contains  $H$ ; (b)  $H \trianglelefteq N_G(H)$ ; (c) for any  $L \leq G$  such that  $H \trianglelefteq L$ ,  $L \leq N_G(H)$ ; (d)  $H \trianglelefteq G$  iff  $N_G(H) = G$ .

**9.230.** Let  $X$  be a  $G$ -set. Prove: for  $g \in G$  and  $x \in X$ ,  $\text{Stab}(gx) = g\text{Stab}(x)g^{-1}$ .

**9.231.** Let  $H$  and  $K$  be subgroups of a group  $G$ . Prove that the  $G$ -sets  $G/H$  and  $G/K$  are isomorphic (as defined in 9.129) iff  $H$  and  $K$  are conjugate subgroups of  $G$  (i.e.,  $K = gHg^{-1}$  for some  $g \in G$ ).

**9.232.** Calculate  $z_\mu$  for every  $\mu \in \text{Par}(6)$ .

**9.233.** Explicitly write down all elements in the centralizer of  $g = (2, 4, 7)(1, 6)(3, 8)(5) \in S_8$ . How large is this centralizer? How large is the conjugacy class of  $g$ ?

**9.234.** Suppose  $f = (2, 4, 7)(8, 10, 15)(1, 9)(11, 12)(17, 20)(18, 19)$  and  $g = (7, 8, 9)(1, 4, 5)(11, 20)(2, 6)(3, 18)(13, 19)$ . How many  $h \in S_{20}$  satisfy  $h \circ f = g \circ h$ ?

**9.235.** Find all integer partitions  $\mu$  of  $n$  for which  $z_\mu = n!$ . Use your answer to calculate  $Z(S_n)$  for all  $n \geq 1$ .

**9.236.** Prove that for all  $n \geq 1$ ,  $p(n) = \frac{1}{n!} \sum_{f \in S_n} z_{\text{type}(f)}$ .

**9.237. Conjugacy Classes of  $A_n$ .** For  $f \in A_n$ , write  $[f]_{A_n}$  (resp.  $[f]_{S_n}$ ) to denote the conjugacy class of  $f$  in  $A_n$  (resp.  $S_n$ ). (a) Show that  $[f]_{A_n} \subseteq [f]_{S_n}$  for all  $f \in A_n$ . (b) Prove: for all  $f \in A_n$ , if there exists  $g \in S_n \sim A_n$  with  $fg = gf$ , then  $[f]_{A_n} = [f]_{S_n}$ ; but if no such  $g$  exists, then  $[f]_{S_n}$  is the disjoint union of  $[f]_{A_n}$  and  $[(1, 2) \circ f \circ (1, 2)]_{A_n}$ , and the latter two conjugacy classes are equal in size. (c) What are the conjugacy classes of  $A_5$ ? How large are they? Use this to prove that  $A_5$  is *simple*, i.e., the only normal subgroups of  $A_5$  are  $\{\text{id}\}$  and  $A_5$ .

**9.238.** Suppose  $G$  is a finite group and  $p$  is a prime divisor of  $|G|$ . Show that the number of elements in  $G$  of order  $p$  is congruent to  $-1 \pmod{p}$ .

**9.239.** (a) Compute  $\binom{8936}{5833} \pmod{7}$ . (b) Compute  $\binom{843}{212} \pmod{10}$ .

**9.240.** Prove 9.138 without using Lucas' congruence, by counting powers of  $p$  in the numerator and denominator of  $\binom{p^a b}{p^a} = (p^a b)_{p^a} / (p^a)!$ .

**9.241. Class Equation.** Let  $G$  be a finite group with center  $Z(G)$  (see 9.125), and let  $x_1, \dots, x_k \in G$  be such that each conjugacy class of  $G$  of size greater than 1 contains exactly one  $x_i$ . Prove that  $|G| = |Z(G)| + \sum_{i=1}^k [G : C_G(x_i)]$ , where each term in the sum is a divisor of  $|G|$  greater than 1.

**9.242.** A  $p$ -group is a finite group of size  $p^e$  for some  $e \geq 1$ . Prove that every  $p$ -group  $G$  has  $|Z(G)| > 1$ .

**9.243. Wilson's Theorem.** Use group actions to prove that if an integer  $p > 1$  is prime, then  $(p-1)! \equiv -1 \pmod{p}$ . Is the converse true?

**9.244.** How many ways are there to color an  $n \times n$  chessboard with  $q$  possible colors if: (a) no symmetries are allowed; (b) rotations of a given board are considered equivalent; (c) rotations and reflections of a given board are considered equivalent?

**9.245.** Consider an  $m \times n$  chessboard where  $m \neq n$ . (a) Describe all symmetries of this board. (b) How many ways can we color such a board with  $q$  possible colors?

**9.246.** How many  $n$ -letter words can be made using a  $k$ -letter alphabet if we identify each word with its reversal?

**9.247.** Consider necklaces that can use  $q$  kinds of gemstones, where rotations and reflections of a given necklace are considered equivalent. How many such necklaces are there with: (a) eight stones; (b) nine stones; (c)  $n$  stones?

**9.248.** Taking rotational symmetries into account, how many ways can we color the vertices of a regular tetrahedron with 7 available colors?

**9.249.** Taking rotational symmetries into account, how many ways can we color the vertices of a cube with 8 available colors?

**9.250.** Taking rotational symmetries into account, how many ways can we color the faces of a cube with  $q$  available colors?

**9.251.** Taking rotational symmetries into account, how many ways can we color the edges of a cube with  $q$  available colors?

**9.252.** Taking all symmetries into account, how many ways are there to color the vertices of the cycle  $C_3$  with three *distinct* colors chosen from a set of five colors?

**9.253.** Taking all symmetries into account, how many ways are there to color the vertices of the cycle  $C_6$  so that three vertices are blue, two are red, and one is yellow?

**9.254.** Taking rotational symmetries into account, how many ways are there to color the vertices of a regular tetrahedron so that: (a) two are blue and two are red; (b) one is red, one is blue, one is green, and one is yellow?

**9.255.** Taking rotational symmetries into account, how many ways are there to color the vertices of a cube so that four are blue, two are red, and two are green?

**9.256.** Taking rotational symmetries into account, how many ways are there to color the faces of a cube so that: (a) three are red, two are blue, and one is green; (b) two are red, two are blue, one is green, and one is yellow?

**9.257.** Taking rotational symmetries into account, how many ways are there to color the edges of a cube so that four are red, four are blue, and four are yellow?

**9.258.** How many ways can we color a  $4 \times 4$  chessboard with five colors (identifying rotations of a given board) if each color must be used at least once?

**9.259.** How many ways can we build an eight-stone necklace using five kinds of gems (identifying rotations and reflections of a given necklace) if each type of gem must be used at least once?

---

## Notes

For a more detailed development of group theory, we recommend the excellent book by Rotman [119]. More information on groups, rings, and fields may be found in textbooks on abstract algebra such as [29, 70, 71]. Many facts about matrices and determinants, including the Cauchy-Binet formula, appear in the matrix theory text by Lancaster [82]. The proof of Cauchy's theorem given in 9.136 is due to McKay [91]. The proof of Lucas' congruence in 9.137 is due to Sagan [120]. The proof of Sylow's theorem given in 9.139 is usually attributed to Wielandt [137], although Miller [92] gave a proof in a similar spirit over 40 years earlier. Proofs of Fermat's little theorem and Wilson's theorem using group actions were given by Peterson [103].

This page intentionally left blank

## Tableaux and Symmetric Polynomials

In this chapter, we study combinatorial objects called *tableaux*. Informally, a tableau is a filling of the cells in the diagram of an integer partition with labels that may be subject to certain ordering conditions. We use tableaux to give a combinatorial definition of *Schur polynomials*, which are examples of *symmetric polynomials*. The theory of symmetric polynomials nicely demonstrates the interplay between combinatorics and algebra. We give a brief introduction to this vast subject in this chapter, stressing bijective proofs throughout.

### 10.1 Partition Diagrams and Skew Shapes

The reader may find it helpful at this point to review the basic definitions concerning integer partitions (see §2.8). Table 10.1 summarizes the notation used in this chapter to discuss integer partitions. In combinatorial arguments, we usually visualize the diagram  $\text{dg}(\mu)$  as a collection of unit boxes, where  $(i, j) \in \text{dg}(\mu)$  corresponds to the box in row  $i$  and column  $j$ . The conjugate partition  $\mu'$  is the partition whose diagram is obtained from  $\text{dg}(\mu)$  by interchanging the roles of rows and columns.

Before defining tableaux, we need the notion of a *skew shape*.

**10.1. Definition: Skew Shapes.** Let  $\mu$  and  $\nu$  be two integer partitions such that  $\text{dg}(\nu) \subseteq \text{dg}(\mu)$ , or equivalently,  $\nu_i \leq \mu_i$  for all  $i \geq 1$ . In this situation, we define the *skew shape*

$$\mu/\nu = \text{dg}(\mu) \sim \text{dg}(\nu) = \{(i, j) : 1 \leq i \leq \ell(\mu), \nu_i < j \leq \mu_i\}.$$

We can visualize  $\mu/\nu$  as the collection of unit squares obtained by starting with the diagram of  $\mu$  and erasing the squares in the diagram of  $\nu$ . If  $\nu = 0 = (0, 0, \dots)$  is the zero

**TABLE 10.1**

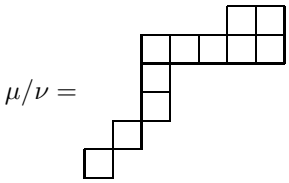
Notation related to integer partitions.

Notation	Definition
$\text{Par}(k)$	set of integer partitions of $k$
$p(k)$	number of integer partitions of $k$
$\text{Par}_N(k)$	set of integer partitions of $k$ with at most $N$ parts
$\mu \vdash k$	$\mu$ is an integer partition of $k$
$\mu_i$	the $i$ th largest part of the partition $\mu$
$\ell(\mu)$	the number of nonzero parts of the partition $\mu$
$\text{dg}(\mu)$	the diagram of $\mu$ , i.e., $\{(i, j) \in \mathbb{N} \times \mathbb{N} : 1 \leq i \leq \ell(\mu), 1 \leq j \leq \mu_i\}$
$\mu'$	conjugate partition to $\mu$
$(1^{a_1} 2^{a_2} \dots k^{a_k} \dots)$	the partition with $a_k$ parts equal to $k$ for $k \geq 1$

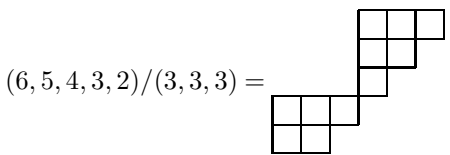


partition, then  $\mu/0 = \text{dg}(\mu)$ . A skew shape of the form  $\mu/0$  is sometimes called a *straight shape*.

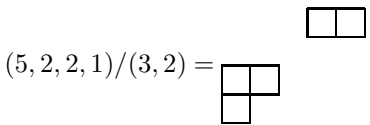
**10.2. Example.** If  $\mu = (7, 7, 3, 3, 2, 1)$  and  $\nu = (5, 2, 2, 2, 1)$ , then



Similarly,



Skew shapes need not be connected; for instance,



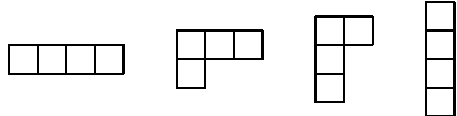
The skew shape  $\mu/\nu$  does not always determine  $\mu$  and  $\nu$  uniquely; for example,

$$(5, 2, 2, 1)/(3, 2) = (5, 3, 2, 1)/(3, 3).$$

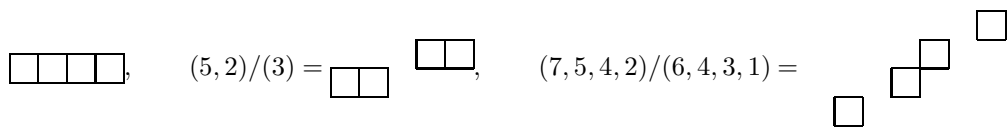
Some special skew shapes will arise frequently in the sequel.

**10.3. Definition: Hooks and Strips.** A *hook* is a skew shape of the form  $(a, 1^{n-a})/0$  for some  $a \leq n$ . A *horizontal strip* is a skew shape that contains at most one cell in each column. A *vertical strip* is a skew shape that contains at most one cell in each row.

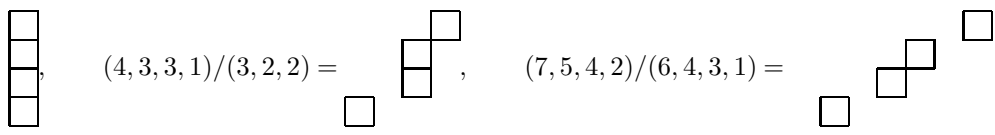
**10.4. Example.** The following picture displays the four hooks of size 4.



The following skew shapes are horizontal strips of size 4.



The following skew shapes are vertical strips of size 4.



## 10.2 Tableaux

Now we are ready to define tableaux.

**10.5. Definition: Tableaux.** Let  $\mu/\nu$  be a skew shape, and let  $X$  be a set. A *tableau of shape  $\mu/\nu$  with values in the alphabet  $X$*  is a function  $T : \mu/\nu \rightarrow X$ .

Informally, we obtain a tableau from the skew shape  $\mu/\nu$  by filling each box  $c \in \mu/\nu$  with a letter  $T(c) \in X$ . We often take  $\nu = 0$ , in which case  $T$  is called a *tableau of shape  $\mu$* . Note that the plural form of “tableau” is “tableaux”; both words are pronounced *tab-loh*.

**10.6. Example.** The following picture displays a tableau  $T$  of shape  $(5, 5, 2)$  with values in  $\mathbb{N}$ :

4	3	3	7	2
1	1	3	9	1
5	6			

Formally,  $T$  is the function with domain  $\text{dg}((5, 5, 2))$  such that

$$T((1, 1)) = 4, \quad T((1, 2)) = T((1, 3)) = 3, \quad \dots, \quad T((3, 1)) = 5, \quad T((3, 2)) = 6.$$

As another example, here is a tableau of shape  $(2, 2, 2, 2)$  with values in  $\{a, b, c, d\}$ :

a	b
b	d
d	a
c	c

Here is a tableau of shape  $(3, 3, 3)/(2, 1)$  with values in  $\mathbb{Z}$ :

		-1
	0	7
4	4	-1

In most discussions of tableaux, we take the alphabet to be either  $\{1, 2, \dots, N\}$  for some fixed  $N$ , or  $\mathbb{N}^+ = \{1, 2, 3, \dots\}$ , or  $\mathbb{Z}$ .

**10.7. Definition: Semistandard Tableaux and Standard Tableaux.** Let  $T$  be a tableau of shape  $\mu/\nu$  taking values in an *ordered* set  $(X, \leq)$ .  $T$  is *semistandard* iff  $T((i, j)) \leq T((i, j + 1))$  for all  $i, j$  such that  $(i, j)$  and  $(i, j + 1)$  both belong to  $\mu/\nu$ ; and  $T((i, j)) < T((i + 1, j))$  for all  $i, j$  such that  $(i, j)$  and  $(i + 1, j)$  both belong to  $\mu/\nu$ . A *standard* tableau is a bijection  $T : \mu/\nu \rightarrow \{1, 2, \dots, n\}$  that is also a semistandard tableau, where  $n = |\mu/\nu|$ .

Less formally, a tableau  $T$  is semistandard iff the entries in each row of  $T$  weakly increase from left to right, and the entries in each column of  $T$  strictly increase from top to bottom. A semistandard tableau is standard iff it contains each number from 1 to  $n$  exactly once. The alphabet  $X$  is usually a subset of  $\mathbb{Z}$  with the usual ordering. Semistandard tableaux are sometimes called *Young tableaux* (in honor of Alfred Young, one of the pioneers in the subject) or *column-strict tableaux*.

**10.8. Example.** Consider the following three tableaux of shape  $(3, 2, 2)$ :

$$T_1 = \begin{array}{|c|c|c|} \hline 1 & 1 & 3 \\ \hline 3 & 4 & \\ \hline 5 & 5 & \\ \hline \end{array} \quad T_2 = \begin{array}{|c|c|c|} \hline 1 & 2 & 6 \\ \hline 3 & 5 & \\ \hline 4 & 7 & \\ \hline \end{array} \quad T_3 = \begin{array}{|c|c|c|} \hline 1 & 2 & 5 \\ \hline 3 & 2 & \\ \hline 4 & 5 & \\ \hline \end{array}.$$

$T_1$  is semistandard but not standard;  $T_2$  is both standard and semistandard;  $T_3$  is neither standard nor semistandard.  $T_3$  is not semistandard because of the strict decrease  $3 > 2$  in row 2, and also because of the weak increase  $2 \leq 2$  in column 2.

**10.9. Example.** There are five standard tableaux of shape  $(3, 2)$ , namely:

$$S_1 = \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 4 & 5 & \\ \hline \end{array} \quad S_2 = \begin{array}{|c|c|c|} \hline 1 & 2 & 4 \\ \hline 3 & 5 & \\ \hline \end{array} \quad S_3 = \begin{array}{|c|c|c|} \hline 1 & 2 & 5 \\ \hline 3 & 4 & \\ \hline \end{array} \quad S_4 = \begin{array}{|c|c|c|} \hline 1 & 3 & 4 \\ \hline 2 & 5 & \\ \hline \end{array} \quad S_5 = \begin{array}{|c|c|c|} \hline 1 & 3 & 5 \\ \hline 2 & 4 & \\ \hline \end{array}.$$

As mentioned in the introduction, there is an amazing formula for counting the number of standard tableaux of a given shape  $\mu$ . This formula is proved in §12.10.

**10.10. Example.** Here is a semistandard tableau of shape  $(6, 5, 5, 3)/(3, 2, 2)$ :

$$S = \begin{array}{ccccccc} & & & & 1 & 1 & 4 \\ & & & 2 & 3 & 4 & \\ & & 4 & 4 & 5 & & \\ 3 & 7 & 7 & & & & \end{array}$$

It will be convenient to have notation for certain sets of tableaux.

**10.11. Definition:**  $\text{SSYT}_X(\mu/\nu)$  and  $\text{SYT}(\mu/\nu)$ . For every skew shape  $\mu/\nu$  and every ordered alphabet  $X$ , let  $\text{SSYT}_X(\mu/\nu)$  be the set of all semistandard tableaux with shape  $\mu/\nu$  taking values in  $X$ . When  $X = \{1, 2, \dots, N\}$ , we abbreviate the notation to  $\text{SSYT}_N(\mu/\nu)$ . Let  $\text{SYT}(\mu/\nu)$  be the set of all standard tableaux of shape  $\mu/\nu$ .

If  $\nu = 0$ , then we omit it from the notation; for instance, 10.9 displays the five elements in the set  $\text{SYT}((3, 2))$ . Observe that  $\text{SYT}(\mu/\nu)$  is a finite set of tableaux. On the other hand,  $\text{SSYT}_X(\mu/\nu)$  is finite iff  $X$  is finite.

### 10.3 Schur Polynomials

We now introduce a weight function on tableaux that keeps track of the number of times each label is used.

**10.12. Definition: Content of a Tableau.** Let  $T$  be a tableau of shape  $\mu/\nu$  with values in  $\mathbb{N}^+$ . The *content* of  $T$  is the infinite sequence  $c(T) = (c_1, c_2, \dots)$ , where  $c_k$  is the number of times the label  $k$  appears in  $T$ . Formally,  $c_k = |\{(i, j) \in \text{dg}(\mu) : T((i, j)) = k\}|$ . Every  $c_k$  is a nonnegative integer, and the sum of all  $c_k$ 's is  $|\mu/\nu|$ . Given variables (indeterminates)  $x_1, x_2, \dots$ , the *content monomial* of  $T$  is

$$x^{c(T)} = x_1^{c_1} x_2^{c_2} \cdots x_k^{c_k} \cdots = \prod_{u \in \mu/\nu} x_{T(u)}.$$

**10.13. Example.** Consider the tableaux from 10.8. The content of  $T_1$  is  $c(T_1) = (2, 0, 2, 1, 2, 0, 0, \dots)$ , and the content monomial of  $T_1$  is  $x^{c(T_1)} = x_1^2 x_3^2 x_4 x_5^2$ . Similarly,

$$x^{c(T_2)} = x_1 x_2 x_3 x_4 x_5 x_6 x_7, \quad x^{c(T_3)} = x_1 x_2^2 x_3 x_4 x_5^2.$$

All five standard tableaux in 10.9 have content monomial  $x_1 x_2 x_3 x_4 x_5$ . More generally, the content monomial of any  $S \in \text{SYT}(\mu/\nu)$  will be  $\prod_{i=1}^n x_i$ , where  $n = |\mu/\nu|$ . The tableau  $S$  shown in 10.10 has content  $c(S) = (2, 1, 2, 4, 1, 0, 2, 0, 0, \dots)$ .

We can now define the Schur polynomials, which are essentially generating functions for semistandard tableaux weighted by content. Recall (§7.16) that  $\mathbb{Q}[x_1, \dots, x_N]$  is the ring of all formal polynomials in the variables  $x_1, \dots, x_N$  with rational coefficients.

**10.14. Definition: Schur Polynomials and Skew Schur Polynomials.** For each integer  $N \geq 1$  and every integer partition  $\mu$ , define the *Schur polynomial for  $\mu$  in  $N$  variables* by setting

$$s_\mu(x_1, \dots, x_N) = \sum_{T \in \text{SSYT}_N(\mu)} x^{c(T)}.$$

More generally, for any skew shape  $\mu/\nu$ , define the *skew Schur polynomial for  $\mu/\nu$*  by setting

$$s_{\mu/\nu}(x_1, \dots, x_N) = \sum_{T \in \text{SSYT}_N(\mu/\nu)} x^{c(T)}.$$

**10.15. Example.** Let us compute the Schur polynomials  $s_\mu(x_1, x_2, x_3)$  for all partitions of 3. First, when  $\mu = (3)$ , we have the following semistandard tableaux of shape (3) using the alphabet  $\{1, 2, 3\}$ :

$$\begin{array}{cccccc} \boxed{1}\boxed{1}\boxed{1} & \boxed{1}\boxed{1}\boxed{2} & \boxed{1}\boxed{1}\boxed{3} & \boxed{1}\boxed{2}\boxed{2} & \boxed{1}\boxed{2}\boxed{3} \\ \boxed{1}\boxed{3}\boxed{3} & \boxed{2}\boxed{2}\boxed{2} & \boxed{2}\boxed{2}\boxed{3} & \boxed{2}\boxed{3}\boxed{3} & \boxed{3}\boxed{3}\boxed{3} \end{array}$$

It follows that

$$s_{(3)}(x_1, x_2, x_3) = x_1^3 + x_1^2x_2 + x_1^2x_3 + x_1x_2^2 + x_1x_2x_3 + x_1x_3^2 + x_2^3 + x_2^2x_3 + x_2x_3^2 + x_3^3.$$

Second, when  $\mu = (2, 1)$ , we obtain the following semistandard tableaux:

$$\begin{array}{ccccccccc} \boxed{1}\boxed{1} & \boxed{1}\boxed{1} & \boxed{2}\boxed{2} & \boxed{1}\boxed{2} & \boxed{1}\boxed{2} & \boxed{1}\boxed{3} & \boxed{1}\boxed{3} & \boxed{2}\boxed{3} \\ \boxed{2} & \boxed{3} & \boxed{3} & \boxed{2} & \boxed{3} & \boxed{2} & \boxed{3} & \boxed{3} \end{array}$$

So  $s_{(2,1)}(x_1, x_2, x_3) = x_1^2x_2 + x_1^2x_3 + x_2^2x_3 + x_1x_2^2 + 2x_1x_2x_3 + x_1x_3^2 + x_2x_3^2$ . Third, when  $\mu = (1, 1, 1)$ , we see that  $s_{(1,1,1)}(x_1, x_2, x_3) = x_1x_2x_3$ , since there is only one semistandard tableau in this case.

Now consider what happens when we change the number of variables. Suppose first that we use  $N = 2$  instead of  $N = 3$ . This means that the allowed alphabet for the tableaux has changed to  $\{1, 2\}$ . Consulting the tableaux just computed, but disregarding those that use the letter 3, we conclude that

$$s_{(3)}(x_1, x_2) = x_1^3 + x_1^2x_2 + x_1x_2^2 + x_2^3; \quad s_{(2,1)}(x_1, x_2) = x_1^2x_2 + x_1x_2^2; \quad s_{(1,1,1)}(x_1, x_2) = 0.$$

In these examples, note that we can obtain the polynomial  $s_\mu(x_1, x_2)$  from  $s_\mu(x_1, x_2, x_3)$  by setting  $x_3 = 0$ . More generally, we claim that for any  $\mu$  and any  $N > M$ , we can obtain  $s_\mu(x_1, \dots, x_M)$  from  $s_\mu(x_1, \dots, x_N)$  by setting the last  $N - M$  variables equal to zero. To verify this, consider the defining formula

$$s_\mu(x_1, x_2, \dots, x_N) = \sum_{T \in \text{SSYT}_N(\mu)} x^{c(T)}.$$

Upon setting  $x_{M+1} = \dots = x_N = 0$  in this formula, the terms coming from tableaux  $T$  that use letters larger than  $M$  will become zero. We are left with the sum over  $T \in \text{SSYT}_M(\mu)$ , which is precisely  $s_\mu(x_1, x_2, \dots, x_M)$ .

Suppose instead that we increase the number of variables from  $N = 3$  to  $N = 5$ . Here

we must draw new tableaux to find the new Schur polynomial. For instance, the tableaux for  $\mu = (1, 1, 1)$  are:

1	1	1	1	1	1	2	2	2	3
2	2	2	3	3	4	3	3	4	4
3	4	5	4	5	5	4	5	5	5

Accordingly,

$$s_{(1,1,1)}(x_1, x_2, x_3, x_4, x_5) = x_1x_2x_3 + x_1x_2x_4 + x_1x_2x_5 + \cdots + x_3x_4x_5.$$

**10.16. Example.** A semistandard tableau of shape  $(1^k)$  using the alphabet  $X = \{1, 2, \dots, N\}$  is essentially a strictly increasing sequence of  $k$  elements of  $X$ , which can be identified with a  $k$ -element subset of  $X$ . Combining this remark with the definition of Schur polynomials, we conclude that

$$s_{(1^k)}(x_1, \dots, x_N) = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq N} x_{i_1} x_{i_2} \cdots x_{i_k}. \quad (10.1)$$

Similarly, a semistandard tableau of shape  $(k)$  is a weakly increasing sequence of  $k$  elements of  $X$ , which can be identified with a  $k$ -element multiset using letters in  $X$ . So

$$s_{(k)}(x_1, \dots, x_N) = \sum_{1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq N} x_{i_1} x_{i_2} \cdots x_{i_k}. \quad (10.2)$$

**10.17. Example.** Given any integer  $N \geq 4$ , what is the coefficient of  $x_1^2 x_2^2 x_3 x_4$  in the skew Schur polynomial  $s_{(4,3)/(1)}(x_1, \dots, x_N)$ ? The answer is the number of semistandard tableaux of shape  $(4, 3)/(1)$  using exactly two 1's, two 2's, one 3, and one 4. Equivalently, we seek the semistandard tableaux with content  $(2, 2, 1, 1)$ . The required tableaux are shown here:

1	1	2				1	1	3				1	2	2		1	2	3		1	2	4
2	3	4				2	2	4				1	3	4		1	2	4		1	2	3

So the desired coefficient is 6. Next, what is the coefficient of  $x_1 x_2 x_3^2 x_4^2$ ? Now we must find the tableaux of content  $(1, 1, 2, 2)$ , which are the following:

	1	2	3				1	2	4				1	3	3		1	3	4		2	3	3	2	3	4	
3	4	4					3	3	4				2	4	4		2	3	4		1	4	4		1	3	4

Again there are six tableaux, so the coefficient of  $x_1 x_2 x_3^2 x_4^2$  is 6. Finally, what is the coefficient of  $x_1^2 x_2 x_3 x_4^2$ ? Drawing the tableaux of content  $(2, 1, 1, 2)$  produces the following list:

1	1	2				1	1	3				1	1	4		1	2	4		1	3	4		1	2	3
3	4	4				2	4	4				2	3	4		1	3	4		1	2	4		1	4	4

The coefficient is 6 again! One may check that for any rearrangement of the vector  $(2, 1, 1, 2, 0, 0, \dots)$ , the number of semistandard tableaux of shape  $(4, 3)/(1)$  having this content is always 6. This is not a coincidence; it is a consequence of the fact that Schur polynomials are *symmetric*, which we will prove shortly (§10.6).

**10.18. Remark.** We have presented a combinatorial definition of Schur polynomials using semistandard tableaux. One can also define Schur polynomials algebraically as a quotient of two determinants; see 11.45. Alternatively, one can define Schur polynomials using determinants involving the elementary or homogeneous symmetric polynomials to be defined below; see 11.60 and 11.61. Many properties of Schur polynomials can be established either combinatorially or algebraically. In this text, we prefer to give the combinatorial proofs.

## 10.4 Symmetric Polynomials

The examples of Schur polynomials computed in the last section were all *symmetric*; in other words, permuting the subscripts of the  $x$ -variables in any fashion did not change the answer. This section begins our examination of the general theory of symmetric polynomials. Throughout the discussion, we assume that  $K$  is a field containing  $\mathbb{Q}$  (for instance,  $\mathbb{Q}$  or  $\mathbb{R}$  or  $\mathbb{C}$ ), and  $N$  is a fixed positive integer. We will be working in the polynomial ring  $R = K[x_1, \dots, x_N]$  consisting of all polynomials in  $N$  variables with coefficients in  $K$  (see §7.16).

One property of polynomial rings such as  $R$  is that we can substitute arbitrary ring elements for each of the variables  $x_i$ . A formal statement of this “universal mapping property” of  $R$  was given in 7.102. Suppose that  $\sigma \in S_N$  is a given permutation of the subscripts of the  $x$ -variables. According to 7.102, there is a unique ring homomorphism  $E : R \rightarrow R$  such that  $E(c) = c$  for all  $c \in K$  and  $E(x_i) = x_{\sigma(i)}$  for all  $i$ . For any polynomial  $f \in R$ , we often denote  $E(f)$  by  $f(x_{\sigma(1)}, \dots, x_{\sigma(N)})$ . Note, in particular, that  $f(x_1, \dots, x_n) = f$ . Informally, we compute  $E(f)$  by starting with a symbolic expression for  $f$  as a sum of products of  $x_i$ ’s, and then replacing each symbol  $x_i$  by  $x_{\sigma(i)}$ . With this notation in hand, we can now give the formal definition of symmetric polynomials.

**10.19. Definition: Symmetric Polynomials.** A polynomial  $f \in K[x_1, \dots, x_N]$  is *symmetric* iff

$$f(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(N)}) = f(x_1, \dots, x_N) \quad \text{for all } \sigma \in S_N.$$

In other words, any permutation of the variables  $x_i$  leaves  $f$  unchanged. Since any permutation can be achieved by a finite sequence of basic transpositions (see 9.29),  $f$  is symmetric iff for every  $i < N$ , interchanging  $x_i$  and  $x_{i+1}$  in  $f$  leaves  $f$  unchanged.

We now introduce special names for some commonly used symmetric polynomials.

**10.20. Definition: Power Sums.** For every  $k \geq 1$ , the polynomial

$$p_k(x_1, x_2, \dots, x_N) = x_1^k + x_2^k + \dots + x_N^k$$

is evidently symmetric. This polynomial is called the  $k$ th *power-sum* in  $N$  variables.

For example,  $p_3(x_1, x_2, x_3, x_4, x_5) = x_1^3 + x_2^3 + x_3^3 + x_4^3 + x_5^3$ .

**10.21. Definition: Elementary Symmetric Polynomials.** For fixed  $k$  with  $1 \leq k \leq N$ , define the polynomial

$$e_k(x_1, x_2, \dots, x_N) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq N} x_{i_1} x_{i_2} \dots x_{i_k}.$$

The polynomial  $e_k$  is called an *elementary symmetric polynomial* in  $N$  variables. One may check that  $e_k$  is indeed symmetric. We also set  $e_0(x_1, \dots, x_N) = 1$  and  $e_k(x_1, \dots, x_N) = 0$  for all  $k > N$ .

For example,  $e_2(x_1, x_2, x_3, x_4) = x_1 x_2 + x_1 x_3 + x_1 x_4 + x_2 x_3 + x_2 x_4 + x_3 x_4$ . By formula (10.1), we see that  $e_k(x_1, \dots, x_N) = s_{(1^k)}(x_1, \dots, x_N)$ , so that elementary symmetric polynomials are special cases of Schur polynomials.

**10.22. Definition: Complete Symmetric Polynomials.** For fixed  $k \geq 1$ , define the polynomial

$$h_k(x_1, x_2, \dots, x_N) = \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq N} x_{i_1} x_{i_2} \dots x_{i_k}.$$

One may verify that  $h_k$  really is symmetric. We also set  $h_0(x_1, \dots, x_N) = 1$ . The polynomials  $h_k$  are called *complete homogeneous symmetric polynomials* in  $N$  variables.

We call  $h_k$  “complete” because it is the sum of *all* monomials of degree  $k$  in the given variables. For example,  $h_2(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3$ . By formula (10.2), we see that  $h_k(x_1, \dots, x_N) = s_{(k)}(x_1, \dots, x_N)$ , so that complete symmetric polynomials are also special cases of Schur polynomials.

The polynomial  $q(x_1, \dots, x_N) = \sum_{i \neq j} x_i^2 x_j^3$  is readily seen to be symmetric. This example can be generalized as follows.

**10.23. Definition: Monomial Symmetric Polynomials.** Let  $\mu$  be an integer partition with at most  $N$  nonzero parts. Write  $\mu = (\mu_1 \geq \mu_2 \geq \dots \geq \mu_N) \in \mathbb{N}^N$  by adding zero parts if necessary. For any  $\alpha \in \mathbb{N}^N$ , write  $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_N^{\alpha_N}$ . Let  $\text{sort}(\alpha) \in \mathbb{N}^N$  be the unique partition obtained by sorting the entries of  $\alpha$  into weakly decreasing order. Next, let  $M(\mu) = \{\alpha \in \mathbb{N}^N : \text{sort}(\alpha) = \mu\}$ . Finally, define the *monomial symmetric polynomial indexed by  $\mu$*  to be

$$m_\mu(x_1, \dots, x_N) = \sum_{\alpha \in M(\mu)} x^\alpha.$$

Informally,  $m_\mu(x_1, \dots, x_N)$  is the sum of all distinct monomials  $x_1^{\alpha_1} \dots x_N^{\alpha_N}$  whose exponent vector can be rearranged to give  $\mu$ . In this notation, the polynomial  $q$  above is  $m_{(3,2)}(x_1, \dots, x_N)$ . Some of our previous examples are instances of monomial symmetric polynomials. Namely, we have  $p_k(x_1, \dots, x_N) = m_{(k)}(x_1, \dots, x_N)$  and  $e_k(x_1, \dots, x_N) = m_{(1^k)}(x_1, \dots, x_N)$ .

Let us check that  $m_\mu$  really is symmetric. Given  $\sigma \in S_N$ , we have

$$m_\mu(x_{\sigma(1)}, \dots, x_{\sigma(N)}) = \sum_{\alpha \in M(\mu)} x_{\sigma(1)}^{\alpha_1} \dots x_{\sigma(N)}^{\alpha_N} = \sum_{\alpha \in M(\mu)} \prod_{i=1}^N x_{\sigma(i)}^{\alpha_i} = \sum_{\alpha \in M(\mu)} \prod_{j=1}^N x_j^{\alpha_{\sigma^{-1}(j)}}.$$

The last step follows by setting  $j = \sigma(i)$  and rearranging the order of factors in each product. To continue, introduce a new summation variable  $\beta = (\alpha_{\sigma^{-1}(1)}, \dots, \alpha_{\sigma^{-1}(N)})$ . The entries of  $\beta$  are obtained by rearranging the entries of  $\alpha$ , so  $\text{sort}(\beta) = \text{sort}(\alpha) = \mu$ . In fact, the map  $\alpha \mapsto \beta$  is a bijection of  $M(\mu)$  to itself with inverse  $\beta \mapsto (\beta_{\sigma(1)}, \dots, \beta_{\sigma(N)})$ . Since addition is commutative, we can continue the calculation by writing

$$\sum_{\alpha \in M(\mu)} \prod_{j=1}^N x_j^{\alpha_{\sigma^{-1}(j)}} = \sum_{\beta \in M(\mu)} \prod_{j=1}^N x_j^{\beta_j} = m_\mu(x_1, \dots, x_N).$$

**10.24. Definition:  $\Lambda_N$ .** Let  $\Lambda_N$  be the set of all symmetric polynomials in  $K[x_1, \dots, x_N]$ .

If two polynomials  $f$  and  $g$  are symmetric, so are  $f + g$ ,  $-f$ , and  $fg$ . For example,

$$\begin{aligned} (fg)(x_{\sigma(1)}, \dots, x_{\sigma(N)}) &= f(x_{\sigma(1)}, \dots, x_{\sigma(N)})g(x_{\sigma(1)}, \dots, x_{\sigma(N)}) \\ &= f(x_1, \dots, x_N)g(x_1, \dots, x_N) \\ &= (fg)(x_1, \dots, x_N) \quad (\text{for all } \sigma \in S_N). \end{aligned}$$

Also, any constant polynomial  $c \in K$  is certainly symmetric, as is any scalar multiple  $cf$  of a symmetric polynomial  $f$ . These comments imply that  $\Lambda_N$  is a subring and  $K$ -vector subspace of  $K[x_1, \dots, x_N]$ . In particular,  $\Lambda_N$  is a commutative ring with identity and a vector space over  $K$ .

We have just seen that  $\Lambda_N$  is closed under products. So, we can multiply together polynomials of the form  $e_k$ ,  $h_k$ , or  $p_k$  to obtain even more examples of symmetric polynomials. This leads to the following definition.

**10.25. Definition: The Symmetric Polynomials  $e_\alpha$ ,  $h_\alpha$ , and  $p_\alpha$ .** Let  $\alpha = (\alpha_1, \dots, \alpha_s)$  be any sequence of positive integers. Define

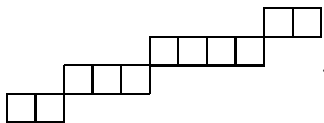
$$\begin{aligned} e_\alpha(x_1, \dots, x_N) &= \prod_{i=1}^s e_{\alpha_i}(x_1, \dots, x_N); \\ h_\alpha(x_1, \dots, x_N) &= \prod_{i=1}^s h_{\alpha_i}(x_1, \dots, x_N); \\ p_\alpha(x_1, \dots, x_N) &= \prod_{i=1}^s p_{\alpha_i}(x_1, \dots, x_N). \end{aligned}$$

We call  $e_\alpha$  the *elementary symmetric polynomial* indexed by  $\alpha$ ;  $h_\alpha$  the *complete homogeneous symmetric polynomial* indexed by  $\alpha$ ; and  $p_\alpha$  the *power-sum symmetric polynomial* indexed by  $\alpha$  (in  $N$  variables).

These definitions are most frequently used when  $\alpha$  is an *integer partition*. Suppose the sequence  $\alpha$  can be sorted to give the partition  $\mu$ . Then  $e_\alpha = e_\mu$ ,  $h_\alpha = h_\mu$ , and  $p_\alpha = p_\mu$ , because multiplication of polynomials is commutative. More generally, if  $\alpha$  and  $\beta$  are rearrangements of each other, then  $e_\alpha = e_\beta$ ,  $h_\alpha = h_\beta$ , and  $p_\alpha = p_\beta$ .

**10.26. Remark.** The power-sum polynomials  $p_\alpha$  have already appeared in our discussion of Pólya's Formula (§9.19), where they were used to count weighted colorings with symmetries taken into account.

**10.27. Remark.** The polynomials  $e_\alpha$  and  $h_\alpha$  are special cases of skew Schur polynomials. For example, consider  $h_\alpha = h_{\alpha_1} h_{\alpha_2} \cdots h_{\alpha_s}$ . We have seen that each factor  $h_{\alpha_i}$  is the generating function for semistandard tableaux of shape  $(\alpha_i)$ . There exists a skew shape  $\mu/\nu$  consisting of disconnected horizontal rows of lengths  $\alpha_1, \dots, \alpha_s$ . When building a semistandard tableau of this shape, each row can be filled with labels independently of the others. So the product rule for weighted sets shows that  $h_\alpha(x_1, \dots, x_N) = s_{\mu/\nu}(x_1, \dots, x_N)$ . For example, given  $h_{(2,4,3,2)} = h_2 h_4 h_3 h_2 = h_{(4,3,2,2)}$ , we draw the skew shape



Thus we have  $h_{(2,4,3,2)} = s_{(11,9,5,2)/(9,5,2)}$ . An analogous procedure works for the  $e_\alpha$ 's, but now we use disconnected vertical columns of lengths given by  $\alpha$ . For example,  $e_{(3,3,1)} = s_{\mu/\nu}$  if we take

$$\mu/\nu = \begin{array}{c} \square \\ \square \\ \square \\ \square \\ \square \\ \square \\ \square \end{array} = (3, 3, 3, 2, 2, 2, 1) / (2, 2, 2, 1, 1, 1).$$

## 10.5 Homogeneous Symmetric Polynomials

When studying symmetric polynomials, it is often helpful to focus attention on those polynomials that are *homogeneous* of a given degree.



**10.28. Definition: Homogeneous Symmetric Polynomials.** For all  $k, N \in \mathbb{N}$ , let  $\Lambda_N^k$  be the set of symmetric polynomials  $p \in \Lambda_N$  such that  $p$  is homogeneous of degree  $k$ . This means that every monomial  $x^\alpha$  appearing in  $p$  with nonzero coefficient has degree  $k$  (i.e.,  $\sum_{i=1}^N \alpha_i = k$ ). In particular, the zero polynomial is homogeneous of every degree.

One can check that for  $f, g \in \Lambda_N^k$  and  $c \in K$ , we have  $f + g \in \Lambda_N^k$  and  $cf \in \Lambda_N^k$ . This means that  $\Lambda_N^k$  is a  $K$ -vector space. Furthermore, the  $K$ -vector space  $\Lambda_N$  is the direct sum of vector spaces

$$\Lambda_N = \bigoplus_{k \geq 0} \Lambda_N^k,$$

since every symmetric polynomial can be uniquely written as a finite sum of its nonzero homogeneous components. Moreover,  $p \in \Lambda_N^k$  and  $q \in \Lambda_N^j$  imply  $pq \in \Lambda_N^{k+j}$ , which means that this direct-sum decomposition turns  $\Lambda_N$  into a *graded ring*.

The vector space  $\Lambda_N$  is infinite-dimensional, but each homogeneous piece  $\Lambda_N^k$  is finite-dimensional. A key theme in the theory of symmetric polynomials is the problem of finding different bases of the vector space  $\Lambda_N^k$  and understanding the relations between these bases. We begin in this section by considering the most straightforward basis for this vector space, which consists of suitable monomial symmetric polynomials.

**10.29. Theorem: Monomial Basis of  $\Lambda_N^k$ .** For every  $k$  and  $N$ , the indexed set of polynomials

$$\{m_\mu(x_1, \dots, x_N) : \mu \in \text{Par}_N(k)\} \subseteq K[x_1, \dots, x_N]$$

is a basis for the  $K$ -vector space  $\Lambda_N^k$ .

*Proof.* For  $\mu \in \text{Par}_N(k)$ , recall that  $m_\mu$  is the sum of all distinct monomials  $x^\alpha$  such that  $\alpha \in \mathbb{N}^N$  can be rearranged to give  $\mu$ . Each of these monomials has degree  $|\mu| = k$ , so that each  $m_\mu$  in the given set is indeed symmetric and homogeneous of degree  $k$ . Next, let us prove that the  $m_\mu$ 's are linearly independent over  $K$ . Suppose some linear combination of these polynomials is the zero polynomial, say

$$\sum_{\mu} c_{\mu} m_{\mu}(x_1, \dots, x_N) = 0 \quad (c_{\mu} \in K). \quad (10.3)$$

Consider some fixed  $\nu \in \text{Par}_N(k)$ . Given any partition  $\mu \neq \nu$ , we cannot rearrange the parts of  $\nu$  to obtain  $\mu$ . It follows that  $m_\nu$  is the only monomial symmetric polynomial in the sum in which  $x^\nu$  appears with nonzero coefficient. The coefficient of  $x^\nu$  in  $m_\nu$  is 1. Extracting the coefficient of  $x^\nu$  on both sides of (10.3) therefore gives  $c_\nu \cdot 1 = 0$ . Since  $\nu$  was arbitrary, all  $c_\nu$ 's are zero, completing the proof of linear independence.

Next, let us prove that the  $m_\mu$ 's span  $\Lambda_N^k$ . Let  $f(x_1, \dots, x_N)$  be any homogeneous symmetric polynomial of degree  $k$ . For each  $\mu \in \text{Par}_N(k)$ , define  $d_\mu \in K$  to be the coefficient of  $x^\mu$  in  $f$ . We claim that

$$\sum_{\mu} d_{\mu} m_{\mu}(x_1, \dots, x_N) = f(x_1, \dots, x_N). \quad (10.4)$$

It suffices to check that, for every  $\alpha \in \mathbb{N}^N$  with  $|\alpha| = k$ , the coefficient of  $x^\alpha$  on both sides of (10.4) is the same. Fix such an  $\alpha$ , and note that there is a unique partition  $\nu \in \text{Par}_N(k)$  such that  $\text{sort}(\alpha) = \nu$ . Reasoning as before, we see that the coefficient of  $x^\alpha$  in  $\sum d_\mu m_\mu$  must be  $d_\nu$ . On the other hand, since  $f$  is symmetric, the coefficient of  $x^\alpha$  in  $f$  must be the same as the coefficient of  $x^\nu$  in  $f$ , since some permutation of the variables will change  $x^\alpha$  into  $x^\nu$ . But the coefficient of  $x^\nu$  in  $f$  is  $d_\nu$  by definition, so we are done.  $\square$

**10.30. Remark.** If  $\mu$  is a partition of  $k$  with more than  $N$  nonzero parts, then  $m_\mu(x_1, \dots, x_N)$  is not defined. If the number of variables  $N$  exceeds  $k$ , then the condition  $\ell(\mu) \leq N$  automatically holds for all partitions  $\mu \vdash k$ . Therefore, when  $N \geq k$ , the basis for  $\Lambda_N^k$  reduces to  $\{m_\mu(x_1, \dots, x_N) : \mu \in \text{Par}(k)\}$ . So, when  $N \geq k$ , we have  $\dim(\Lambda_N^k) = p(k)$ , the number of integer partitions of  $k$ .

## 10.6 Symmetry of Schur Polynomials

Recall the definition of skew Schur polynomials from 10.14:

$$s_{\mu/\nu}(x_1, \dots, x_N) = \sum_{T \in \text{SSYT}_N(\mu/\nu)} x^{c(T)}.$$

We are about to give a bijective proof that the polynomial appearing in this definition is always symmetric. First, we give names to the coefficients of these polynomials.

**10.31. Definition: Kostka Numbers.** For each skew shape  $\mu/\nu$  and each  $\alpha \in \mathbb{N}^N$ , define the *Kostka number*  $K_{\mu/\nu, \alpha}$  to be the coefficient of  $x^\alpha$  in  $s_{\mu/\nu}(x_1, \dots, x_N)$ . Equivalently,  $K_{\mu/\nu, \alpha}$  is the number of semistandard tableaux of shape  $\mu/\nu$  and content  $\alpha$ .

**10.32. Example.** The calculations in 10.17 show that

$$K_{(4,3)/(1),(2,2,1,1)} = K_{(4,3)/(1),(1,1,2,2)} = K_{(4,3)/(1),(2,1,1,2)} = 6.$$

Similarly, we see from 10.9 that  $K_{(3,2),(1,1,1,1,1)} = 5$ .

The following result is the key to proving the symmetry of Schur polynomials.

**10.33. Theorem: Symmetry of Kostka Numbers.** For all skew shapes  $\mu/\nu$  and all  $\alpha, \beta \in \mathbb{N}^N$  such that  $\text{sort}(\alpha) = \text{sort}(\beta)$ , we have  $K_{\mu/\nu, \alpha} = K_{\mu/\nu, \beta}$ .

*Proof.* Fix  $\mu/\nu$  and  $\alpha, \beta$  as in the theorem statement. Since  $\text{sort}(\alpha) = \text{sort}(\beta)$ , we can pass from  $\alpha$  to  $\beta$  by a suitable permutation of the entries of  $\alpha$ . This permutation can be achieved in finitely many steps by repeatedly interchanging two consecutive entries of  $\alpha$  (cf. 9.29 and 9.179). By induction, it therefore suffices to prove the result when  $\beta$  is obtained from  $\alpha$  by switching  $\alpha_i$  and  $\alpha_{i+1}$  for some  $i < N$ .

Let  $Y$  be the set of all tableaux  $T \in \text{SSYT}_N(\mu/\nu)$  such that  $c(T) = \alpha$ , and let  $Z$  be the set of all tableaux  $T \in \text{SSYT}_N(\mu/\nu)$  such that  $c(T) = \beta$ . Since  $|Y| = K_{\mu/\nu, \alpha}$  and  $|Z| = K_{\mu/\nu, \beta}$ , it suffices to define a bijection  $f_i : Y \rightarrow Z$ . The map  $f_i$  must take a semistandard tableau of shape  $\mu/\nu$  and create a new semistandard tableau of the same shape in which the number of  $i$ 's and  $(i+1)$ 's are switched, while the number of  $k$ 's (for all  $k \neq i, i+1$ ) is unchanged. We will illustrate the action of  $f_3$  on the following tableau:

			1	1	1	1	2	3
		2	3	3	3	4	4	4
1	3	3	4	4	5	5	6	
4	4	6	6	6	7	9		
5	6	7	7	8				

Observe that certain occurrences of 3 are “matched” with an occurrence of 4 in the cell

directly below. Let us underline the 3's and 4's that are *not* part of these matched pairs:

			1	1	1	1	2	3
		2	3	3	<u>3</u>	<u>4</u>	<u>4</u>	4
1	3	<u>3</u>	4	4	5	5	6	
<u>4</u>	4	6	6	6	7	9		
<u>5</u>	6	7	7	8				

Notice that each row of the tableau contains a (possibly empty) run of consecutive cells consisting of underlined 3's and 4's. The entries directly above these cells are  $< 3$ , while the entries directly below are  $> 4$ . So we are free to change the frequency of 3's and 4's within each run without affecting the semistandardness of the tableau. If the run in a given row consists of  $j$  threes followed by  $k$  fours (where  $j, k \geq 0$ ), we will change this to a run consisting of  $k$  threes followed by  $j$  fours. Doing this in every row will switch the frequency of 3's and 4's (note that the matched pairs are not touched, and these contribute equally to the frequency counts for 3 and 4). Our example tableau is mapped by  $f_3$  to the following tableau:

			1	1	1	1	2	3
		2	3	3	<u>3</u>	<u>3</u>	<u>4</u>	4
1	3	<u>4</u>	4	4	5	5	6	
<u>3</u>	4	6	6	6	7	9		
<u>5</u>	6	7	7	8				

Applying the same run-modification process to this new tableau will restore the original tableau; this means that  $f_3$  is a bijection. As another example of the action of  $f_3$ , we have

$$f_3 \left( \begin{array}{|c|c|c|c|c|c|c|c|c|} \hline 3 & 3 & 3 & 4 & 4 & 4 & 4 & 4 & 6 \\ \hline 4 & & & & & & & & \\ \hline \end{array} \right) = \begin{array}{|c|c|c|c|c|c|c|c|c|} \hline 3 & 3 & 3 & 3 & 3 & 3 & 4 & 4 & 6 \\ \hline 4 & & & & & & & & \\ \hline \end{array}.$$

The definition of  $f_i$  for general  $i$  is exactly the same. We locate and ignore matched pairs consisting of an  $i$  directly atop an  $i + 1$ , then underline the remaining  $i$ 's and  $(i + 1)$ 's, then switch the relative frequencies of the underlined  $i$ 's and  $(i + 1)$ 's in each row. This action is reversible, maintains semistandardness, and switches the overall frequency of  $i$ 's and  $(i + 1)$ 's while preserving the frequency of all other letters. So we have found the required bijection.  $\square$

**10.34. Example.** The preceding proof allows us to construct explicit bijections between the collections of tableaux in 10.17, which are counted by various Kostka numbers  $K_{(4,3)/(1),\alpha}$  such that  $\text{sort}(\alpha) = (2, 2, 1, 1)$ . As directed by the proof, we must chain together suitable maps  $f_i$ , where the values of  $i$  are chosen to rearrange the starting content vector  $\alpha$  into the target content vector  $\beta$ . For example, we can go from  $(2, 2, 1, 1)$  to  $(2, 1, 2, 1)$  to  $(2, 1, 1, 2)$  by applying  $f_2$  and then  $f_3$ . So, for instance, the first tableau of content  $(2, 2, 1, 1)$  in 10.17 is mapped to a tableau of content  $(2, 1, 1, 2)$  as follows:

$$\begin{array}{|c|c|c|c|} \hline 1 & 1 & 1 & 2 \\ \hline 2 & 3 & 4 & \\ \hline \end{array} \xrightarrow{f_2} \begin{array}{|c|c|c|c|} \hline 1 & 1 & 3 & \\ \hline 2 & 3 & 4 & \\ \hline \end{array} \xrightarrow{f_3} \begin{array}{|c|c|c|c|} \hline 1 & 1 & 4 & \\ \hline 2 & 3 & 4 & \\ \hline \end{array}.$$

If we continue by applying the maps  $f_1$  and then  $f_2$ , we reach a tableau with content  $(1, 1, 2, 2)$ :

$$\begin{array}{|c|c|c|c|} \hline 1 & 1 & 4 & \\ \hline 2 & 3 & 4 & \\ \hline \end{array} \xrightarrow{f_1} \begin{array}{|c|c|c|c|} \hline 2 & 2 & 4 & \\ \hline 1 & 3 & 4 & \\ \hline \end{array} \xrightarrow{f_2} \begin{array}{|c|c|c|c|} \hline 2 & 3 & 4 & \\ \hline 1 & 3 & 4 & \\ \hline \end{array}.$$

The inverse bijection is computed by applying the maps in the reverse order. For example, the first tableau of content  $(1, 1, 2, 2)$  in 10.17 is mapped to a tableau of content  $(2, 2, 1, 1)$  via the following steps.

$$\begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & \\ \hline 3 & 4 & 4 & \\ \hline \end{array} \xrightarrow{f_2} \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & \\ \hline 2 & 4 & 4 & \\ \hline \end{array} \xrightarrow{f_1} \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & \\ \hline 1 & 4 & 4 & \\ \hline \end{array} \xrightarrow{f_3} \begin{array}{|c|c|c|c|} \hline 1 & 2 & 4 & \\ \hline 1 & 3 & 3 & \\ \hline \end{array} \xrightarrow{f_2} \begin{array}{|c|c|c|c|} \hline 1 & 2 & 4 & \\ \hline 1 & 2 & 3 & \\ \hline \end{array}.$$

We can now deduce the symmetry of skew Schur polynomials. In fact, we can even expand these polynomials as linear combinations of monomial symmetric polynomials using the Kostka numbers.

**10.35. Theorem: Monomial Expansion of Schur Polynomials.** For all skew shapes  $\mu/\nu$  with  $k$  boxes and all  $N \geq 1$ ,

$$s_{\mu/\nu}(x_1, \dots, x_N) = \sum_{\rho \in \text{Par}_N(k)} K_{\mu/\nu, \rho} m_{\rho}(x_1, \dots, x_N). \quad (10.5)$$

In particular,  $s_{\mu/\nu}(x_1, \dots, x_N)$  is a *homogeneous, symmetric* polynomial of degree  $k$ .

*Proof.* Consider the following calculation:

$$\begin{aligned} s_{\mu/\nu}(x_1, \dots, x_N) &= \sum_{\alpha \in \mathbb{N}^N} K_{\mu/\nu, \alpha} x^{\alpha} = \sum_{\rho \in \text{Par}_N(k)} \sum_{\substack{\alpha \in \mathbb{N}^N: \\ \text{sort}(\alpha) = \rho}} K_{\mu/\nu, \alpha} x^{\alpha} \\ &= \sum_{\rho \in \text{Par}_N(k)} \sum_{\substack{\alpha \in \mathbb{N}^N: \\ \text{sort}(\alpha) = \rho}} K_{\mu/\nu, \rho} x^{\alpha} = \sum_{\rho \in \text{Par}_N(k)} K_{\mu/\nu, \rho} \sum_{\substack{\alpha \in \mathbb{N}^N: \\ \text{sort}(\alpha) = \rho}} x^{\alpha} \\ &= \sum_{\rho \in \text{Par}_N(k)} K_{\mu/\nu, \rho} m_{\rho}(x_1, \dots, x_N). \end{aligned}$$

The first step follows from the definition of Kostka numbers. In the second step, we reorder the sum by classifying  $\alpha \in \mathbb{N}^N$  based on which partition  $\alpha$  sorts to. Only partitions of  $k$  occur, since  $x^{c(T)}$  is a monomial of degree  $k$  for every tableau  $T$  on the  $k$ -box shape  $\mu/\nu$ . The third step follows from 10.33. The fourth step uses the fact that  $K_{\mu/\nu, \rho}$  does not depend on the inner summation index  $\alpha$ . The final step follows by definition of  $m_{\rho}$ .  $\square$

## 10.7 Orderings on Partitions

We will use 10.35 to find bases for the vector spaces  $\Lambda_N^k$  consisting of suitable Schur polynomials. First, however, we need to introduce some ordering relations on sets of integer partitions.

**10.36. Definition: Lexicographic Ordering of Partitions.** Suppose  $\mu = (\mu_i : i \geq 1)$  and  $\nu = (\nu_i : i \geq 1)$  are partitions of the same integer  $k$ . We say that  $\nu$  is *lexicographically greater* than  $\mu$ , written  $\mu \leq_{\text{lex}} \nu$ , iff either  $\mu = \nu$  or the first nonzero entry in the vector  $\nu - \mu$  is positive. The latter condition means that for some  $j$ ,  $\mu_1 = \nu_1$ ,  $\mu_2 = \nu_2$ ,  $\dots$ ,  $\mu_{j-1} = \nu_{j-1}$ , and  $\mu_j < \nu_j$ .

It is routine to check that  $\leq_{\text{lex}}$  is a *total order* on  $\text{Par}(k)$ , for each  $k \geq 0$ .

**10.37. Example.** Here is a list of all integer partitions of 6, written in lexicographic order from smallest to largest:

$$\begin{aligned} (1, 1, 1, 1, 1, 1) &\leq_{\text{lex}} (2, 1, 1, 1, 1) \leq_{\text{lex}} (2, 2, 1, 1) \leq_{\text{lex}} (2, 2, 2) \leq_{\text{lex}} (3, 1, 1, 1) \\ &\leq_{\text{lex}} (3, 2, 1) \leq_{\text{lex}} (3, 3) \leq_{\text{lex}} (4, 1, 1) \leq_{\text{lex}} (4, 2) \leq_{\text{lex}} (5, 1) \leq_{\text{lex}} (6). \end{aligned}$$

For example,  $(3, 1, 1, 1) \leq_{\text{lex}} (3, 2, 1)$  since

$$(3, 2, 1, 0, 0, \dots) - (3, 1, 1, 1, 0, \dots) = (0, 1, 0, -1, 0, \dots)$$

and the earliest nonzero entry in this vector is positive.

In the coming sections, we will frequently be considering matrices and vectors whose rows and columns are indexed by integer partitions. Unless otherwise specified, we will always use the lexicographic ordering of partitions to determine which partition labels each row and column of the matrix. For instance, when  $k = 3$ , a matrix  $A = (c_{\mu,\nu} : \mu, \nu \in \text{Par}(3))$  will be displayed as follows:

$$A = \begin{bmatrix} c_{(1,1,1),(1,1,1)} & c_{(1,1,1),(2,1)} & c_{(1,1,1),(3)} \\ c_{(2,1),(1,1,1)} & c_{(2,1),(2,1)} & c_{(2,1),(3)} \\ c_{(3),(1,1,1)} & c_{(3),(2,1)} & c_{(3),(3)} \end{bmatrix}.$$

Next we consider a partial ordering on partitions that occurs frequently in the theory of symmetric polynomials.

**10.38. Definition: Dominance Ordering on Partitions.** Let  $\mu = (\mu_i : i \geq 1)$  and  $\nu = (\nu_i : i \geq 1)$  be two partitions of the same integer  $k$ . We say that  $\nu$  *dominates*  $\mu$ , written  $\mu \leq \nu$ , iff

$$\mu_1 + \mu_2 + \cdots + \mu_i \leq \nu_1 + \nu_2 + \cdots + \nu_i \quad \text{for all } i \geq 1.$$

Note that  $\mu \not\leq \nu$  iff there exists an  $i \geq 1$  with  $\mu_1 + \cdots + \mu_i > \nu_1 + \cdots + \nu_i$ .

**10.39. Example.** We have  $(2, 2, 1, 1) \leq (4, 2)$  since  $2 \leq 4$ ,  $2 + 2 \leq 4 + 2$ ,  $2 + 2 + 1 \leq 4 + 2 + 0$ , and  $2 + 2 + 1 + 1 \leq 4 + 2 + 0 + 0$ . On the other hand,  $(3, 1, 1, 1) \not\leq (2, 2, 2)$  since  $3 > 2$ , and  $(2, 2, 2) \not\leq (3, 1, 1, 1)$  since  $2 + 2 + 2 > 3 + 1 + 1$ . This example shows that not every pair of partitions is comparable under the dominance relation.

**10.40. Theorem: Dominance Partial Order.** The dominance relation is a partial ordering on  $\text{Par}(k)$ , for every  $k \geq 0$ .

*Proof.* We will show that  $\leq$  is reflexive, antisymmetric, and transitive on  $\text{Par}(k)$ . *Reflexivity:* Given  $\mu \vdash k$ , we have  $\mu_1 + \cdots + \mu_i \leq \mu_1 + \cdots + \mu_i$  for all  $i \geq 1$ . So  $\mu \leq \mu$ . *Antisymmetry:* Suppose  $\mu, \nu \vdash k$ ,  $\mu \leq \nu$ , and  $\nu \leq \mu$ . We know  $\mu_1 + \cdots + \mu_i \leq \nu_1 + \cdots + \nu_i$  and also  $\nu_1 + \cdots + \nu_i \leq \mu_1 + \cdots + \mu_i$  for all  $i$ , hence  $\mu_1 + \cdots + \mu_i = \nu_1 + \cdots + \nu_i$  for all  $i \geq 1$ . In particular, taking  $i = 1$  gives  $\mu_1 = \nu_1$ . For each  $i > 1$ , subtracting the  $(i - 1)$ th equation from the  $i$ th equation shows that  $\mu_i = \nu_i$ . So  $\mu = \nu$ . *Transitivity:* Fix  $\mu, \nu, \rho \vdash k$ , and assume  $\mu \leq \nu \leq \rho$ ; we must prove  $\mu \leq \rho$ . We know  $\mu_1 + \cdots + \mu_i \leq \nu_1 + \cdots + \nu_i$  for all  $i$ , and also  $\nu_1 + \cdots + \nu_i \leq \rho_1 + \cdots + \rho_i$  for all  $i$ . Combining these inequalities yields  $\mu_1 + \cdots + \mu_i \leq \rho_1 + \cdots + \rho_i$  for all  $i$ , so  $\mu \leq \rho$ .  $\square$

One can check that  $\leq$  is a *total* ordering of  $\text{Par}(k)$  iff  $k \leq 5$ .

**10.41. Theorem: Lexicographic vs. Dominance Ordering.** For all  $\mu, \nu \vdash k$ ,  $\mu \leq \nu$  implies  $\mu \leq_{\text{lex}} \nu$ .

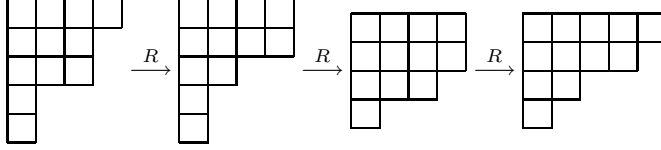
*Proof.* Fix  $\mu, \nu$  such that  $\mu \not\leq_{\text{lex}} \nu$ ; we will prove that  $\mu \not\leq \nu$ . By definition of the lexicographic order, there must exist an index  $j \geq 1$  such that  $\mu_i = \nu_i$  for all  $i < j$ , but  $\mu_j > \nu_j$ . Adding these relations together, we see that  $\mu_1 + \cdots + \mu_j > \nu_1 + \cdots + \nu_j$ , and so  $\mu \not\leq \nu$ .  $\square$

The next definition and theorem will allow us to visualize the dominance relation in terms of partition diagrams.

**10.42. Definition: Raising Operation.** Let  $\mu$  and  $\nu$  be two partitions of  $k$ . We say that  $\nu$  is related to  $\mu$  by a *raising operation*, denoted  $\mu R \nu$ , iff there exist  $i < j$  such that  $\nu_i = \mu_i + 1$ ,  $\nu_j = \mu_j - 1$ , and  $\nu_s = \mu_s$  for all  $s \neq i, j$ .

Intuitively,  $\mu R \nu$  means that we can go from the diagram for  $\mu$  to the diagram for  $\nu$  by taking the last square from some row of  $\mu$  and moving it to the end of a higher row.

**10.43. Example.** The following pictures illustrate a sequence of raising operations.



Observe that  $(4, 3, 3, 1, 1) \trianglelefteq (5, 4, 2, 1)$ , so that the last partition in the sequence dominates the first one. The next result shows that this always happens.

**10.44. Theorem: Dominance and Raising Operations.** Given  $\mu, \nu \vdash k$ , we have  $\mu \trianglelefteq \nu$  iff there exist  $m \geq 0$  and partitions  $\mu^0, \dots, \mu^m$  such that  $\mu = \mu^0 R \mu^1 R \mu^2 \cdots \mu^{m-1} R \mu^m = \nu$ .

*Proof.* Let us first show that  $\mu R \nu$  implies  $\mu \trianglelefteq \nu$ . Suppose  $\nu = (\mu_1, \dots, \mu_i + 1, \dots, \mu_j - 1, \dots)$  as in the definition of dominance ordering. Let us check that  $\mu_1 + \cdots + \mu_k \leq \nu_1 + \cdots + \nu_k$  holds for all  $k \geq 1$ . This is true for  $k < i$ , since equality holds for these  $k$ 's. If  $k = i$ , note that  $\nu_1 + \cdots + \nu_k = \mu_1 + \cdots + \mu_{i-1} + (\mu_i + 1)$ , so the required inequality does hold. Similarly, for all  $k$  with  $i \leq k < j$ , we have  $\nu_1 + \cdots + \nu_k = \mu_1 + \cdots + \mu_k + 1 > \mu_1 + \cdots + \mu_k$ . Finally, for all  $k \geq j$ , we have  $\mu_1 + \cdots + \mu_k = \nu_1 + \cdots + \nu_k$  since the  $+1$  and  $-1$  adjustments to parts  $i$  and  $j$  cancel out.

Next, suppose  $\mu$  and  $\nu$  are linked by a chain of raising operations as in the theorem statement, say  $\mu = \mu^0 R \mu^1 R \mu^2 \cdots R \mu^m = \nu$ . By what has just been proved, we have  $\mu = \mu^0 \trianglelefteq \mu^1 \trianglelefteq \mu^2 \cdots \trianglelefteq \mu^m = \nu$ . Since  $\trianglelefteq$  is transitive, we conclude that  $\mu \trianglelefteq \nu$ , as desired.

Conversely, suppose that  $\mu \trianglelefteq \nu$ . Consider the vector  $(d_1, d_2, \dots)$  such that  $d_s = (\nu_1 + \cdots + \nu_s) - (\mu_1 + \cdots + \mu_s)$ . Since  $\mu \trianglelefteq \nu$ , we have  $d_s \geq 0$  for all  $s$ . Also,  $d_s = 0$  for all large enough  $s$  since  $\mu$  and  $\nu$  are both partitions of  $k$ . We argue by induction on  $n = \sum_s d_s$ . If  $n = 0$ , then  $\mu = \nu$ , and we can take  $m = 0$  and  $\mu = \mu^0 = \nu$ . Otherwise, let  $i$  be the least index such that  $d_i > 0$ , and let  $j$  be the least index after  $i$  such that  $d_j = 0$ . The choice of  $i$  shows that  $\mu_s = \nu_s$  for all  $s < i$ , but  $\mu_i < \nu_i$ . If  $i > 1$ , the inequality  $\mu_i < \nu_i \leq \nu_{i-1} = \mu_{i-1}$  shows that it is possible to add one box to the end of row  $i$  in  $\text{dg}(\mu)$  and still get a partition diagram. If  $i = 1$ , the addition of this box will certainly give a partition diagram. On the other hand, the relations  $d_{j-1} > 0$ ,  $d_j = 0$  mean that  $\mu_1 + \cdots + \mu_{j-1} < \nu_1 + \cdots + \nu_{j-1}$  but  $\mu_1 + \cdots + \mu_j = \nu_1 + \cdots + \nu_j$ , so that  $\mu_j > \nu_j$ . Furthermore, from  $d_j = 0$  and  $d_{j+1} \geq 0$  we deduce that  $\mu_{j+1} \leq \nu_{j+1}$ . So,  $\mu_{j+1} \leq \nu_{j+1} \leq \nu_j < \mu_j$ , which shows that we can remove a box from row  $j$  of  $\text{dg}(\mu)$  and still get a partition diagram.

We have just shown that it is permissible to modify  $\mu$  by a raising operator that moves the box at the end of row  $j$  to the end of row  $i$ . Let  $\mu^1$  be the new partition obtained in this way, so that  $\mu R \mu^1$ . Consider how the partial sums  $\mu_1 + \cdots + \mu_s$  change when we replace  $\mu$  by  $\mu^1$ . For  $s < i$  or  $s \geq j$ , the partial sums are the same for  $\mu$  and  $\mu^1$ . For  $i \leq s < j$ , the partial sums increase by 1. Since  $d_s > 0$  in the range  $i \leq s < j$ , it follows that the new differences  $d'_s = (\nu_1 + \cdots + \nu_s) - (\mu^1_1 + \cdots + \mu^1_s)$  are all  $\geq 0$ ; in other words,  $\mu^1 \trianglelefteq \nu$ . We have  $d'_s = d_s - 1$  for  $i \leq s < j$ , and  $d'_s = d_s$  for all other  $s$ ; so  $\sum d'_s < \sum d_s$ . Arguing by induction, we can find a chain of raising operations linking  $\mu^1$  to  $\nu$ . This completes the inductive proof.  $\square$

As an application of the previous result, we prove the following fact relating the dominance ordering to the conjugation operation on partitions.

**10.45. Theorem: Dominance vs. Conjugation.** For all  $\mu, \nu \in \text{Par}(k)$ ,  $\mu \trianglelefteq \nu$  iff  $\nu' \trianglelefteq \mu'$ .

*Proof.* Fix  $\mu, \nu \in \text{Par}(k)$ . Note first that  $\mu R \nu$  implies  $\nu' R \mu'$ . This assertion follows from the pictorial description of the raising operation, since the box that moves from a lower row in  $\mu$  to a higher row in  $\nu$  necessarily moves from some column to a column strictly to its

right. Reversing the direction of motion and transposing the diagrams, we see that we can go from  $\nu'$  to  $\mu'$  by moving a box in  $\nu'$  from a lower row to a higher row.

Next, assuming  $\mu \trianglelefteq \nu$ , 10.44 shows that there is a chain

$$\mu = \mu^0 R \mu^1 R \mu^2 \cdots \mu^{m-1} R \mu^m = \nu.$$

Applying the remark in the previous paragraph to each link in this chain gives a new chain

$$\nu' = (\mu^m)' R (\mu^{m-1})' R \cdots (\mu^2)' R (\mu^1)' R (\mu^0)' = \mu'.$$

Invoking 10.44 again, we see that  $\nu' \trianglelefteq \mu'$ .

Conversely, assume that  $\nu' \trianglelefteq \mu'$ . Applying the result just proved, we get  $\mu'' \trianglelefteq \nu''$ . Since  $\mu'' = \mu$  and  $\nu'' = \nu$ , we have  $\mu \trianglelefteq \nu$ .  $\square$

## 10.8 Schur Bases

We now have all the necessary tools to find bases for the vector spaces  $\Lambda_N^k$  consisting of Schur polynomials. First we illustrate the key ideas with an example.

**10.46. Example.** In 10.15, we computed the Schur polynomials  $s_\mu(x_1, x_2, x_3)$  for all partitions  $\mu \in \text{Par}(3)$ . We can use 10.35 to write these Schur polynomials as linear combinations of monomial symmetric polynomials, where the coefficients are Kostka numbers:

$$\begin{aligned} s_{(1,1,1)}(x_1, x_2, x_3) &= m_{(1,1,1)}(x_1, x_2, x_3); \\ s_{(2,1)}(x_1, x_2, x_3) &= 2m_{(1,1,1)}(x_1, x_2, x_3) + m_{(2,1)}(x_1, x_2, x_3); \\ s_{(3)}(x_1, x_2, x_3) &= m_{(1,1,1)}(x_1, x_2, x_3) + m_{(2,1)}(x_1, x_2, x_3) + m_{(3)}(x_1, x_2, x_3). \end{aligned}$$

These equations can be combined to give the following matrix identity:

$$\begin{bmatrix} s_{(1,1,1)} \\ s_{(2,1)} \\ s_{(3)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} m_{(1,1,1)} \\ m_{(2,1)} \\ m_{(3)} \end{bmatrix}.$$

The  $3 \times 3$  matrix appearing here is lower-triangular with ones on the main diagonal, hence is invertible. Multiplying by the inverse matrix, we find that

$$\begin{bmatrix} m_{(1,1,1)} \\ m_{(2,1)} \\ m_{(3)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} s_{(1,1,1)} \\ s_{(2,1)} \\ s_{(3)} \end{bmatrix}.$$

This says that each monomial symmetric polynomial  $m_\nu(x_1, x_2, x_3)$  is expressible as a linear combination of the Schur polynomials  $s_\mu(x_1, x_2, x_3)$ . Since the  $m_\nu$ 's form a basis of the vector space  $\Lambda_3^3$ , the Schur polynomials must span this space. Since  $\dim(\Lambda_3^3) = p(3) = 3$ , the three-element set  $\{s_\mu(x_1, x_2, x_3) : \mu \vdash 3\}$  is in fact a basis of  $\Lambda_3^3$ .

The argument given in the example extends to the general case. The key fact is that the transition matrix from Schur polynomials to monomial symmetric polynomials is always lower-triangular with ones on the main diagonal, as shown next.

**10.47. Theorem: Lower Unitriangularity of the Kostka Matrix.** For all partitions  $\lambda$ ,  $K_{\lambda, \lambda} = 1$ . For all partitions  $\lambda$  and  $\mu$ ,  $K_{\lambda, \mu} \neq 0$  implies  $\mu \trianglelefteq \lambda$  (and also  $\mu \leq_{\text{lex}} \lambda$ , by 10.41).

*Proof.* The Kostka number  $K_{\lambda,\lambda}$  is the number of semistandard tableaux  $T$  of shape  $\lambda$  and content  $\lambda$ . Such a tableau must contain  $\lambda_i$  copies of  $i$  for each  $i \geq 1$ . In particular,  $T$  contains  $\lambda_1$  ones. Since  $T$  is semistandard, all these ones must occur in the top row, which has  $\lambda_1$  boxes. So the top row of  $T$  contains all ones. For the same reason, the  $\lambda_2$  twos in  $T$  must all occur in the second row, which has  $\lambda_2$  boxes. Arguing similarly, we see that  $T$  must be the tableau whose  $i$ th row contains all  $i$ 's, for  $i \geq 1$ . Thus, there is exactly one semistandard tableaux of shape  $\lambda$  and content  $\lambda$ . For example, when  $\lambda = (4, 2, 2, 1)$ , we must have

$$T = \begin{array}{|c|c|c|c|} \hline 1 & 1 & 1 & 1 \\ \hline 2 & 2 & & \\ \hline 3 & 3 & & \\ \hline 4 & & & \\ \hline \end{array}.$$

For the second part of the theorem, we argue by contradiction. Assume  $\lambda, \mu \in \text{Par}(k)$  are such that  $K_{\lambda,\mu} \neq 0$  and yet  $\mu \not\leq \lambda$ . Since the Kostka number is nonzero, there exists a semistandard tableau  $T$  of shape  $\lambda$  and content  $\mu$ . Since the content of  $T$  is  $\mu$ , the entries of  $T$  come from the alphabet  $\{1, 2, \dots, \ell(\mu)\}$ . Since the columns of  $T$  must strictly increase, we observe that all 1's in  $T$  must occur in row 1; all 2's in  $T$  must occur in row 1 or row 2; and, in general, all  $j$ 's in  $T$  must occur in the top  $j$  rows of  $\text{dg}(\lambda)$ . Now, the assumption  $\mu \not\leq \lambda$  means that there is an  $i \geq 1$  with  $\mu_1 + \dots + \mu_i > \lambda_1 + \dots + \lambda_i$ . The left side of this inequality is the total number of occurrences of the symbols  $1, 2, \dots, i$  in  $T$ . The right side of the inequality is the total number of boxes in the top  $i$  rows of  $T$ . Our preceding observation now produces the desired contradiction, since there is not enough room in the top  $i$  rows of  $\text{dg}(\lambda)$  to accommodate the  $\mu_1 + \dots + \mu_i$  occurrences of the symbols  $1, 2, \dots, i$  in  $T$ .  $\square$

**10.48. Example.** Let  $\lambda = (3, 2, 2)$  and  $\mu = (2, 2, 2, 1)$ . The Kostka number  $K_{\lambda,\mu}$  is 3, as we see by listing the semistandard tableaux of shape  $\lambda$  and content  $\mu$ :

$$\begin{array}{|c|c|c|} \hline 1 & 1 & 2 \\ \hline 2 & 3 & \\ \hline 3 & 4 & \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline 1 & 1 & 3 \\ \hline 2 & 2 & \\ \hline 3 & 4 & \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline 1 & 1 & 4 \\ \hline 2 & 2 & \\ \hline 3 & 3 & \\ \hline \end{array}$$

In each tableau, all occurrences of  $i$  appear in the top  $i$  rows, and we do have  $\mu \leq \lambda$ .

**10.49. Theorem: Schur Basis of  $\Lambda_N^k$ .** For all  $k, N \in \mathbb{N}$ , the set of Schur polynomials

$$\{s_\lambda(x_1, \dots, x_N) : \lambda \in \text{Par}_N(k)\} \subseteq K[x_1, \dots, x_N]$$

is a basis of the  $K$ -vector space  $\Lambda_N^k$ .

*Proof.* Let  $p = |\text{Par}_N(k)|$ , and let  $\mathbf{S}$  be the  $p \times 1$  column vector consisting of the Schur polynomials  $\{s_\lambda(x_1, \dots, x_N) : \lambda \in \text{Par}_N(k)\}$ , arranged in lexicographic order. Let  $\mathbf{M}$  be the  $p \times 1$  column vector consisting of the monomial symmetric polynomials  $\{m_\mu(x_1, \dots, x_N) : \mu \in \text{Par}_N(k)\}$ , also arranged in lexicographic order. Finally, let  $\mathbf{K}$  be the  $p \times p$  matrix, with rows and columns indexed by elements of  $\text{Par}_N(k)$  in lexicographic order, such that the entry in row  $\lambda$  and column  $\mu$  is the Kostka number  $K_{\lambda,\mu}$ . Now 10.35 says that, for every  $\lambda \in \text{Par}_N(k)$ ,

$$s_\lambda(x_1, \dots, x_N) = \sum_{\mu \in \text{Par}_N(k)} K_{\lambda,\mu} m_\mu(x_1, \dots, x_N).$$

These scalar equations are equivalent to the matrix-vector equation  $\mathbf{S} = \mathbf{KM}$ . Moreover, 10.47 asserts that  $\mathbf{K}$  is a lower-triangular matrix of integers with 1's on the main diagonal. So  $\mathbf{K}$  has an inverse matrix (whose entries are also integers, since  $\det(\mathbf{K}) = 1$ ).



Multiplying on the left by this inverse matrix, we get  $\mathbf{M} = \mathbf{K}^{-1}\mathbf{S}$ . This equation means that every  $m_\mu$  is a linear combination of Schur polynomials. Since the  $m_\mu$ 's generate  $\Lambda_N^k$ , the Schur polynomials must also generate this space. Linear independence follows automatically since the number of Schur polynomials in the proposed basis (namely  $p$ ) equals the dimension of the vector space, by 10.29.  $\square$

**10.50. Remark.** The matrix  $\mathbf{K}$  occurring in this proof is called a *Kostka matrix*. The entries of the *inverse Kostka matrix*  $\mathbf{K}^{-1}$  tell us how to expand monomial symmetric polynomials in terms of Schur polynomials. As seen in the  $3 \times 3$  example, these entries are integers which can be negative. It is natural to ask for a combinatorial interpretation for these matrix entries in terms of suitable collections of *signed* objects. One such interpretation will be discussed in §11.15 below.

**10.51. Remark.** If  $\lambda \in \text{Par}(k)$  has more than  $N$  parts, then  $s_\lambda(x_1, \dots, x_N) = 0$ . This follows since there are not enough letters available in the alphabet to fill the first column of  $\text{dg}(\lambda)$  with a strictly increasing sequence. So there are no semistandard tableaux of this shape on this alphabet.

## 10.9 Tableau Insertion

We have seen that the Kostka numbers give the coefficients of the monomial expansion of Schur polynomials. Remarkably, the Kostka numbers also relate Schur polynomials to the elementary and complete homogeneous symmetric polynomials. This fact will be a consequence of the *Pieri rules*, which tell us how to rewrite products of the form  $s_\mu e_k$  and  $s_\mu h_k$  as linear combinations of Schur polynomials.

To develop these results, we need a fundamental combinatorial construction on tableaux called *tableau insertion*. Given a semistandard tableau  $T$  of shape  $\mu$  and a letter  $x$ , we wish to build a new semistandard tableau by “inserting  $x$  into  $T$ .” The following recursive procedure allows us to do this.

**10.52. Definition: Tableau Insertion Algorithm.** Let  $T$  be a semistandard tableau of straight shape  $\mu$  over the ordered alphabet  $X$ , and let  $x \in X$ . We define a new tableau, denoted  $T \leftarrow x$ , by the following procedure.

1. If  $\mu = 0$ , so that  $T$  is the empty tableau, then  $T \leftarrow x$  is the tableau of shape (1) whose sole entry is  $x$ .
2. Otherwise, let  $y_1 \leq y_2 \leq \dots \leq y_t$  be the entries in the top row of  $T$ .
  - 2a. If  $y_t \leq x$ , then  $T \leftarrow x$  is the tableau of shape  $(\mu_1 + 1, \mu_2, \dots)$  obtained by placing a new box containing  $x$  at the right end of the top row of  $T$ .
  - 2b. Otherwise, choose the minimal  $i \in \{1, 2, \dots, t\}$  such that  $x < y_i$ . Let  $T'$  be the semistandard tableaux consisting of all rows of  $T$  after the first one. To form  $T \leftarrow x$ , first replace  $y_i$  by  $x$  in the top row of  $T$ . Then replace  $T'$  by  $T' \leftarrow y_i$ , which is computed recursively by the same algorithm.

If step 2b occurs, we say that  $x$  has *bumped*  $y_i$  out of row 1. In turn,  $y_i$  may bump an element from row 2 to row 3, and so on.

This recursively defined insertion algorithm always terminates, since the number of times we execute step 2b is at most  $\ell(\mu)$ , which is finite. We must also prove that the algorithm

always produces a tableau that is *semistandard* and of *partition shape*. We will prove these facts after considering some examples.

**10.53. Example.** Let us compute  $T \leftarrow 3$ , where

$$T = \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & 4 & 4 & 6 \\ \hline 2 & 4 & 5 & 6 & 6 & & \\ \hline 3 & 5 & 7 & 8 & & & \\ \hline 4 & 6 & & & & & \\ \hline \end{array}.$$

We scan the top row of  $T$  from left to right, looking for the first entry *strictly larger* than 3. This entry is the 4 in the fifth box. In step 2b, the 3 bumps the 4 into the second row. The current situation looks like this:

$$\begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & \underline{3} & 4 & 6 \\ \hline 2 & 4 & 5 & 6 & 6 & & \\ \hline 3 & 5 & 7 & 8 & & & \\ \hline 4 & 6 & & & & & \\ \hline \end{array} \leftarrow 4$$

Now we scan the second row from left to right, looking for the first entry strictly larger than 4. It is the 5, so the 4 bumps the 5 into the third row:

$$\begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & \underline{3} & 4 & 6 \\ \hline 2 & 4 & \underline{4} & 6 & 6 & & \\ \hline 3 & 5 & 7 & 8 & & & \\ \hline 4 & 6 & & & & & \\ \hline \end{array} \leftarrow 5$$

Next, the 5 bumps the 7 into the fourth row:

$$\begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & \underline{3} & 4 & 6 \\ \hline 2 & 4 & \underline{4} & 6 & 6 & & \\ \hline 3 & 5 & \underline{5} & 8 & & & \\ \hline 4 & 6 & & & & & \\ \hline \end{array} \leftarrow 7$$

Now, everything in the fourth row is weakly smaller than 7. So, as directed by step 2a, we insert 7 at the end of this row. The final tableau is therefore

$$T \leftarrow 3 = \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & \underline{3} & 4 & 6 \\ \hline 2 & 4 & \underline{4} & 6 & 6 & & \\ \hline 3 & 5 & \underline{5} & 8 & & & \\ \hline 4 & 6 & \underline{7} & & & & \\ \hline \end{array}.$$

We have underlined the entries of  $T \leftarrow 3$  that were affected by the insertion process. These entries are the starting value  $x = 3$  together with those entries that got bumped during the insertion. Call these entries the *bumping sequence*; in this example, the bumping sequence is  $(3, 4, 5, 7)$ . The sequence of boxes occupied by the bumping sequence is called the *bumping path*. The lowest box in the bumping path is called the *new box*. It is the only box in  $T \leftarrow 3$  that was not present in the original diagram for  $T$ .

For a simpler example of tableau insertion, note that

$$T \leftarrow 6 = \begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & 4 & 4 & 6 & \underline{6} \\ \hline 2 & 4 & 5 & 6 & 6 & & & \\ \hline 3 & 5 & 7 & 8 & & & & \\ \hline 4 & 6 & & & & & & \\ \hline \end{array}.$$

The reader should check that

$$T \leftarrow 1 = \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & \underline{1} & 3 & 4 & 4 & 6 \\ \hline 2 & 2 & 5 & 6 & 6 & & \\ \hline 3 & 4 & 7 & 8 & & & \\ \hline 4 & \underline{5} & & & & & \\ \hline \underline{6} & & & & & & \\ \hline \end{array}, \quad T \leftarrow 0 = \begin{array}{|c|c|c|c|c|c|c|} \hline \underline{0} & 1 & 2 & 3 & 4 & 4 & 6 \\ \hline \underline{1} & 4 & 5 & 6 & 6 & & \\ \hline 2 & 5 & 7 & 8 & & & \\ \hline \underline{3} & 6 & & & & & \\ \hline \underline{4} & & & & & & \\ \hline \end{array}.$$

To prove that  $T \leftarrow x$  is always semistandard of partition shape, we need the following result.

**10.54. Theorem: Bumping Sequence and Bumping Path.** Given a semistandard tableau  $T$  and element  $x$ , let  $(x_1, x_2, \dots, x_k)$  be the bumping sequence and let  $((1, j_1), (2, j_2), \dots, (k, j_k))$  be the bumping path arising in the computation of  $T \leftarrow x$ . Then  $x = x_1 < x_2 < \dots < x_k$  and  $j_1 \geq j_2 \geq \dots \geq j_k > 0$ . (So the bumping sequence *strictly increases* and the bumping path *moves weakly left* as it goes down.)

*Proof.* By definition of the bumping sequence,  $x = x_1$  and  $x_i$  bumps  $x_{i+1}$  from row  $i$  into row  $i + 1$ , for  $1 \leq i < k$ . By definition of bumping,  $x_i$  bumps an entry strictly larger than itself, so  $x_i < x_{i+1}$  for all  $i < k$ . Next, consider what happens to  $x_{i+1}$  when it is bumped out of row  $i$ . Before being bumped,  $x_{i+1}$  occupied the cell  $(i, j_i)$ . After being bumped,  $x_{i+1}$  will occupy the cell  $(i + 1, j_{i+1})$ , which is either an existing cell in row  $i + 1$  of  $T$ , or a new cell at the end of this row. Consider the cell  $(i + 1, j_i)$  directly below  $(i, j_i)$ . If this cell is outside the shape of  $T$ , the previous observation shows that  $(i + 1, j_{i+1})$  must be located weakly left of this cell, so that  $j_{i+1} \leq j_i$ . On the other hand, if  $(i + 1, j_i)$  is part of  $T$  and contains some value  $z$ , then  $x_{i+1} < z$  because  $T$  is semistandard. Now,  $x_{i+1}$  bumps the *leftmost* entry in row  $i + 1$  that is strictly larger than  $x_{i+1}$ . Since  $z$  is such an entry,  $x_{i+1}$  bumps  $z$  or some entry to the left of  $z$ . In either case, we again have  $j_{i+1} \leq j_i$ .  $\square$

**10.55. Theorem: Output of a Tableau Insertion.** If  $T$  is a semistandard tableau of shape  $\mu$ , then  $T \leftarrow x$  is a semistandard tableau whose shape is a partition obtained by adding one new box to  $\text{dg}(\mu)$ .

*Proof.* Let us first show that the shape of  $T \leftarrow x$  is a partition diagram. This shape is obtained from  $\text{dg}(\mu)$  by adding one new box (the last box in the bumping path). If this new box is in the top row, then the resulting shape is certainly a partition diagram (namely,  $\text{dg}((\mu_1 + 1, \mu_2, \dots))$ ). Suppose the new box is in row  $i > 1$ . Then 10.54 shows that the new box is located weakly left of a box in the previous row that belongs to  $\text{dg}(\mu)$ . This implies that  $\mu_i < \mu_{i-1}$ , so adding the new box to row  $i$  will still give a partition diagram.

Next we prove that each time an entry of  $T$  is bumped during the insertion of  $x$ , the resulting tableau is still semistandard. Suppose, at some stage in the insertion process, that an element  $y$  bumps  $z$  out of the following configuration:

$$\begin{array}{|c|c|c|} \hline & a & \\ \hline b & z & c \\ \hline & d & \\ \hline \end{array}.$$

(Some of the boxes containing  $a, b, c, d$  may be absent, in which case the following argument should be modified appropriately.) The original configuration is part of a semistandard tableau, so  $b \leq z \leq c$  and  $a < z < d$ . Because  $y$  bumps  $z$ ,  $z$  must be the first entry strictly larger than  $y$  in its row. This means that  $b \leq y < z \leq c$ , so replacing  $z$  by  $y$  will still leave a weakly increasing row. Does the column containing  $z$  still strictly increase after the bumping? On one hand,  $y < d$ , since  $y < z < d$ . On the other hand, if the box containing  $a$  exists (i.e., if  $z$  is below the top row), then  $y$  was the element bumped out of  $a$ 's row. Since the bumping path moves weakly left, the original location of  $y$  must have been weakly right of  $z$  in the row above  $z$ . If  $y$  was directly above  $z$ , then  $a$  must have bumped  $y$ , and so  $a < y$  by definition of bumping. Otherwise,  $y$  was located strictly to the right of  $a$  before  $y$  was bumped, so  $a \leq y$ . We cannot have  $a = y$  in this situation, since otherwise  $a$  (or something to its left) would have been bumped instead of  $y$ . Thus,  $a < y$  in all cases.

Finally, consider what happens at the end of the insertion process, when an element  $w$  is inserted in a new box at the end of a (possibly empty) row. This only happens when  $w$

weakly exceeds all entries in its row, so the row containing  $w$  is weakly increasing. There is no cell below  $w$  in this case. Repeating the argument at the end of the last paragraph, we see that  $w$  is strictly greater than the entry directly above it (if any). This completes the proof that  $T \leftarrow x$  is semistandard.  $\square$

## 10.10 Reverse Insertion

Given the output  $T \leftarrow x$  of a tableau insertion operation, it is generally not possible to determine what  $T$  and  $x$  were. However, if we also know the location of the new box created by this insertion, then we can recover  $T$  and  $x$ . More generally, we can start with any semistandard tableau  $S$  and any “corner box” of  $S$ , and “uninsert” the value in this box to obtain a semistandard tableau  $T$  and value  $x$  such that  $S = T \leftarrow x$ . (Here we do not assume in advance that  $S$  has the form  $T \leftarrow x$ .) This process is called *reverse tableau insertion*. Before giving the general definition, we consider some examples.

**10.56. Example.** Consider the following semistandard tableau:

$$S = \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 2 & 2 & 4 \\ \hline 2 & 2 & 3 & 5 & \\ \hline 3 & 4 & 4 & 6 & \\ \hline 4 & 5 & & & \\ \hline 6 & 6 & & & \\ \hline 7 & 8 & & & \\ \hline \end{array}$$

There are three corner boxes whose removal from  $S$  will still leave a partition diagram; they are the boxes at the end of the first, third, and sixth rows. Removing the corner box in the top row, we evidently will have  $S = T_1 \leftarrow 4$ , where

$$T_1 = \begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & 2 \\ \hline 2 & 2 & 3 & 5 \\ \hline 3 & 4 & 4 & 6 \\ \hline 4 & 5 & & \\ \hline 6 & 6 & & \\ \hline 7 & 8 & & \\ \hline \end{array}$$

Suppose instead that we remove the 6 at the end of the third row of  $S$ . Reversing the bumping process, we see that 6 must have been bumped into the third row from the second row. What element bumped it? In this case, it is the 5 in the second row. In turn, the 5 must have originally resided in the first row, before being bumped into the second row by the 4. In summary, we have  $S = T_2 \leftarrow \underline{4}$ , where

$$T_2 = \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 2 & 2 & \underline{5} \\ \hline 2 & 2 & 3 & \underline{6} & \\ \hline 3 & 4 & 4 & & \\ \hline 4 & 5 & & & \\ \hline 6 & 6 & & & \\ \hline 7 & 8 & & & \\ \hline \end{array}$$

(Here we have underlined the entries in the *reverse bumping sequence*, which occupy boxes in the *reverse bumping path*.) Finally, consider what happens when we uninsert the 8 at the end of the last row of  $S$ . The 8 was bumped to its current location by one of the 6’s in the previous row; it must have been bumped by the rightmost 6, lest semistandardness be

violated. Next, the 6 was bumped by the 5 in row 4; the 5 was bumped by the rightmost 4 in row 3; and so on. In general, to determine which element in row  $i$  bumped some value  $z$  into row  $i + 1$ , we look for the *rightmost* entry in row  $i$  that is *strictly less* than  $z$ . Continuing in this way, we discover that  $S = T_3 \leftarrow 2$ , where

$$T_3 = \begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & \underline{3} & 4 \\ \hline 2 & 2 & \underline{4} & 5 & \\ \hline 3 & 4 & \underline{5} & 6 & \\ \hline 4 & \underline{6} & & & \\ \hline 6 & \underline{8} & & & \\ \hline 7 & & & & \\ \hline \end{array}$$

With these examples in hand, we are ready to give the general definition of reverse tableau insertion.

**10.57. Definition: Reverse Tableau Insertion.** Suppose  $S$  is a semistandard tableau of shape  $\nu$ . A *corner box* of  $\nu$  is a box  $(i, j) \in \text{dg}(\nu)$  such that  $\text{dg}(\nu) \sim \{(i, j)\}$  is still the diagram of some partition  $\mu$ . Given  $S$  and a corner box  $(i, j)$  of  $\nu$ , we define a tableau  $T$  and a value  $x$  as follows. We will construct a *reverse bumping sequence*  $(x_i, x_{i-1}, \dots, x_1)$  and a *reverse bumping path*  $((i, j_i), (i-1, j_{i-1}), \dots, (1, j_1))$  as follows.

1. Set  $j_i = j$  and  $x_i = S((i, j))$ , which is the value of  $S$  in the given corner box.
2. Once  $x_k$  and  $j_k$  have been found, for some  $i \geq k > 1$ , scan row  $k-1$  of  $S$  for the rightmost entry that is strictly less than  $x_k$ . Define  $x_{k-1}$  to be this entry, and let  $j_{k-1}$  be the column in which this entry occurs.
3. At the end, let  $x = x_1$ , and let  $T$  be the tableau obtained by erasing box  $(i, j_i)$  from  $S$  and replacing the contents of box  $(k-1, j_{k-1})$  by  $x_k$  for  $i \geq k > 1$ .

The next results will show that reverse insertion really is the two-sided inverse of ordinary insertion (given knowledge of the location of the new box).

**10.58. Theorem: Properties of Reverse Insertion.** Suppose we perform reverse tableau insertion on  $S$  and  $(i, j)$  to obtain  $T$  and  $x$  as in 10.57. (a) The reverse bumping sequence satisfies  $x_i > x_{i-1} > \dots > x_1 = x$ . (b) The reverse bumping path satisfies  $j_i \leq j_{i-1} \leq \dots \leq j_1$ . (c)  $T$  is a semistandard tableau of shape  $\mu$ . (d)  $(T \leftarrow x) = S$ .

*Proof.* Part (a) follows from the definition of  $x_{k-1}$  in 10.57. Note that there does exist an entry in row  $k-1$  strictly less than  $x_k$ , since the entry directly above  $x_k$  (in cell  $(k-1, j_k)$  of  $S$ ) is such an entry. This observation also shows that the rightmost entry strictly less than  $x_k$  in row  $k-1$  occurs in column  $j_k$  or later, proving (b). Part (c) follows from (a) and (b) by an argument similar to that given in 10.55; we let the reader fill in the details. For part (d), consider the bumping sequence  $(x'_1, x'_2, \dots)$  and bumping path  $((1, j'_1), (2, j'_2), \dots)$  for the forward insertion  $T \leftarrow x$ . We have  $x'_1 = x = x_1$  by definition. Recall that  $x_1 = S((1, j_1))$  was the rightmost entry in row 1 of  $S$  that was strictly less than  $x_2$ , and  $T((1, j_1)) = x_2$  by definition of  $T$ . All other entries in row 1 are the same in  $S$  and  $T$ . So  $T((1, j_1)) = x_2$  will be the leftmost entry of row 1 of  $T$  strictly larger than  $x_1$ . So, in the insertion  $T \leftarrow x$ ,  $x_1$  bumps  $x_2$  out of cell  $(1, j_1)$ . In particular,  $j'_1 = j_1$  and  $x'_2 = x_2$ . Repeating this argument in each successive row, we see by induction that  $x'_k = x_k$  and  $j'_k = j_k$  for all  $k$ . At the end of the insertion, we have recovered the starting tableau  $S$ .  $\square$

**10.59. Theorem: Reversing Insertion.** Suppose  $S = (T \leftarrow x)$  for some semistandard tableau  $T$  and value  $x$ . Let  $(i, j)$  be the new box created by this insertion. If we perform reverse insertion on  $S$  starting with box  $(i, j)$ , we will obtain the original  $T$  and  $x$ .

*Proof.* This can be proved by induction, showing step by step that the forward and reverse bumping paths and bumping sequences are the same. The argument is similar to part (d) of 10.58, so we leave it as an exercise for the reader.  $\square$

The next theorem summarizes the results of the last two sections.

**10.60. Theorem: Invertibility of Tableau Insertion.** Let  $X$  be an ordered set, and let  $\mu$  be a fixed partition. Let  $P(\mu)$  be the set of all partitions that can be obtained from  $\mu$  by adding a single box at the end of some row. There exist mutually inverse bijections

$$I : \text{SSYT}_X(\mu) \times X \rightarrow \bigcup_{\nu \in P(\mu)} \text{SSYT}_X(\nu), \quad R : \bigcup_{\nu \in P(\mu)} \text{SSYT}_X(\nu) \rightarrow \text{SSYT}_X(\mu) \times X$$

given by  $I(T, x) = T \leftarrow x$  and  $R(S) =$  the result of applying reverse tableau insertion to  $S$  starting at the unique box of  $S$  not in  $\mu$ .

*Proof.* We have seen that  $I$  and  $R$  are well-defined functions mapping into the stated codomains. We see from 10.58(d) that  $I \circ R$  is the identity map on  $\bigcup_{\nu \in P(\mu)} \text{SSYT}_X(\nu)$ , while 10.59 says that  $R \circ I$  is the identity map on  $\text{SSYT}_X(\mu) \times X$ . Hence  $I$  and  $R$  are bijections.  $\square$

Let us take  $X = \{1, 2, \dots, N\}$  in 10.60. We can regard  $X$  as a weighted set with  $\text{wt}(i) = x_i$ . The generating function for this weighted set is  $x_1 + x_2 + \dots + x_N = h_1(x_1, \dots, x_N) = s_{(1)}(x_1, \dots, x_N) = e_1(x_1, \dots, x_N)$ . Note that the content monomial  $x^{c(T \leftarrow j)}$  is  $x^{c(T)}x_j$ , since  $T \leftarrow j$  contains all the entries of  $T$  together with one new entry equal to  $j$ . This means that  $\text{wt}(I(T, j)) = \text{wt}(T) \text{wt}(j)$ , so that the bijection  $I$  in the theorem is *weight-preserving*. Using the product rule for weighted sets and the definition of Schur polynomials, the generating function for the domain of  $I$  is  $s_\mu(x_1, \dots, x_N)h_1(x_1, \dots, x_N)$ . Using the sum rule for weighted sets, the generating function for the codomain of  $I$  is  $\sum_{\nu \in P(\mu)} s_\nu(x_1, \dots, x_N)$ . To summarize, our tableau insertion algorithms have furnished a combinatorial proof of the following *multiplication rule*:

$$s_\mu h_1 = s_\mu e_1 = s_\mu s_{(1)} = \sum_{\nu \in P(\mu)} s_\nu,$$

where we sum over all partitions  $\nu$  obtained by adding *one* corner box to  $\mu$ . We have discovered the simplest instance of the *Pieri rules* mentioned at the beginning of §10.9.

## 10.11 Bumping Comparison Theorem

We now extend the analysis of the previous section to prove the general Pieri rules for expanding  $s_\mu h_k$  and  $s_\mu e_k$  in terms of Schur polynomials. The key idea is to see what happens when we successively insert  $k$  weakly increasing numbers (or  $k$  strictly decreasing numbers) into a semistandard tableau by repeated tableau insertion. We begin with some examples to build intuition.

**10.61. Example.** Consider the semistandard tableau

$$T = \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & 4 \\ \hline 2 & 3 & 3 & 4 & \\ \hline 3 & 4 & 5 & 6 & \\ \hline 5 & 5 & 6 & 7 & \\ \hline 6 & & & & \\ \hline \end{array}$$

Let us compute the tableaux that result by successively inserting 2, 3, 3, 5 into  $T$ :

$$\begin{aligned}
 T_1 = T \leftarrow 2 &= \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 2 & \underline{2} & 4 \\ \hline 2 & 3 & 3 & \underline{3} & \\ \hline 3 & 4 & \underline{4} & 6 & \\ \hline 5 & 5 & \underline{5} & 7 & \\ \hline 6 & \underline{6} & & & \\ \hline \end{array} ; & T_2 = T_1 \leftarrow 3 &= \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 2 & 2 & \underline{3} \\ \hline 2 & 3 & 3 & 3 & \underline{4} \\ \hline 3 & 4 & 4 & 6 & \\ \hline 5 & 5 & 5 & 7 & \\ \hline 6 & 6 & & & \\ \hline \end{array} ; \\
 T_3 = T_2 \leftarrow 3 &= \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 2 & 2 & 3 & \underline{3} \\ \hline 2 & 3 & 3 & 3 & 4 & \\ \hline 3 & 4 & 4 & 6 & & \\ \hline 5 & 5 & 5 & 7 & & \\ \hline 6 & 6 & & & & \\ \hline \end{array} ; & T_4 = T_3 \leftarrow 5 &= \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 2 & 3 & 3 & \underline{5} \\ \hline 2 & 3 & 3 & 3 & 4 & & \\ \hline 3 & 4 & 4 & 6 & & & \\ \hline 5 & 5 & 5 & 7 & & & \\ \hline 6 & 6 & & & & & \\ \hline \end{array} .
 \end{aligned}$$

Consider the skew shape consisting of the four cells in  $T_4$  that are not in  $T_1$ , which are marked by asterisks in the following picture:

					*	*
					*	
	*					

Observe that this skew shape is a *horizontal strip* of size 4. Next, compare the bumping paths in the successive insertions of 2, 3, 3, 5. We see that each path lies *strictly right* of the previous bumping path and ends with a new box in a *weakly higher* row.

Now return to the original tableau  $T$ , and consider the insertion of a strictly decreasing sequence 5, 4, 2, 1. We obtain the following tableaux:

$$\begin{aligned}
 S_1 = T \leftarrow 5 &= \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & 4 & \underline{5} \\ \hline 2 & 3 & 3 & 4 & & \\ \hline 3 & 4 & 5 & 6 & & \\ \hline 5 & 5 & 6 & 7 & & \\ \hline 6 & & & & & \\ \hline \end{array} ; & S_2 = S_1 \leftarrow 4 &= \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & 4 & \underline{4} \\ \hline 2 & 3 & 3 & 4 & \underline{5} & \\ \hline 3 & 4 & 5 & 6 & & \\ \hline 5 & 5 & 6 & 7 & & \\ \hline 6 & & & & & \\ \hline \end{array} ; \\
 S_3 = S_2 \leftarrow 2 &= \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & \underline{2} & 4 & 4 \\ \hline 2 & 3 & 3 & \underline{3} & 5 & \\ \hline 3 & 4 & 4 & 6 & & \\ \hline 5 & 5 & \underline{5} & 7 & & \\ \hline 6 & \underline{6} & & & & \\ \hline \end{array} ; & S_4 = S_3 \leftarrow 1 &= \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & \underline{1} & 2 & 4 & 4 \\ \hline 2 & \underline{2} & 3 & 3 & 5 & \\ \hline 3 & \underline{3} & 4 & 6 & & \\ \hline 4 & 5 & 5 & 7 & & \\ \hline \underline{5} & 6 & & & & \\ \hline \underline{6} & & & & & \\ \hline \end{array} .
 \end{aligned}$$

This time, each successive bumping path is *weakly left* of the previous one and ends in a *strictly lower* row. Accordingly, the new boxes in  $S_4$  form a *vertical strip*:

					*
					*
	*				
*					

We now show that the observations in this example hold in general.

**10.62. Bumping Comparison Theorem.** Let  $T$  be a semistandard tableau using letters in  $X$ , and let  $x, y \in X$ . Let the new box in  $T \leftarrow x$  be  $(i, j)$ , and let the new box in  $(T \leftarrow x) \leftarrow y$  be  $(r, s)$ . (a)  $x \leq y$  iff  $i \geq r$  and  $j < s$ ; (b)  $x > y$  iff  $i < r$  and  $j \geq s$ .

*Proof.* It suffices to prove the forward implications, since exactly one of  $x \leq y$  or  $x > y$  is true. Let the bumping path for the insertion of  $x$  be  $((1, j_1), (2, j_2), \dots, (i, j_i))$  (where  $j_i = j$ ), and let the bumping sequence be  $(x = x_1, x_2, \dots, x_i)$ . Let the bumping path for the insertion of  $y$  be  $((1, s_1), (2, s_2), \dots, (r, s_r))$  (where  $s_r = s$ ), and let the bumping sequence be  $(y = y_1, y_2, \dots, y_r)$ .

Assume  $x \leq y$ . We prove the following statement by induction: for all  $k$  with  $1 \leq k \leq r$ , we have  $i \geq k$  and  $x_k \leq y_k$  and  $j_k < s_k$ . When  $k = 1$ , we have  $i \geq 1$  and  $x_1 \leq y_1$  (by assumption). Note that  $x_1$  appears in box  $(1, j_1)$  of  $T \leftarrow x$ . We cannot have  $s_1 \leq j_1$ , for this would mean that  $y_1$  bumps an entry weakly left of  $(1, j_1)$ , and this entry is at most  $x_1 \leq y_1$ , contrary to the definition of bumping. So  $j_1 < s_1$ . Now consider the induction step. Suppose  $k < r$  and the induction hypothesis holds for  $k$ ; does it hold for  $k + 1$ ? Since  $k < r$ ,  $y_k$  must have bumped something from position  $(k, s_k)$  into the next row. Since  $j_k < s_k$ ,  $x_k$  must also have bumped something out of row  $k$ , proving that  $i \geq k + 1$ . The object bumped by  $x_k$ , namely  $x_{k+1}$ , appears to the left of the object bumped by  $y_k$ , namely  $y_{k+1}$ , in the same row of a semistandard tableau. Therefore,  $x_{k+1} \leq y_{k+1}$ . Now we can repeat the argument used for the first row to see that  $j_{k+1} < s_{k+1}$ . Now that the induction is complete, take  $k = r$  to see that  $i \geq r$  and  $j = j_i \leq j_r < s_r = s$  (the first inequality holding since the bumping path for  $T \leftarrow x$  moves weakly left as we go down).

Next, assume  $x > y$ . This time we prove the following by induction: for all  $k$  with  $1 \leq k \leq i$ , we have  $r > k$  and  $x_k > y_k$  and  $j_k \geq s_k$ . When  $k = 1$ , we have  $x_1 = x > y = y_1$ . Since  $x$  appears somewhere in the first row of  $T \leftarrow x$ ,  $y$  will necessarily bump something into the second row, so  $r > 1$ . In fact, the thing bumped by  $y$  occurs weakly left of the position  $(1, j_1)$  occupied by  $x$ , so  $s_1 \leq j_1$ . For the induction step, assume the induction hypothesis is known for some  $k < i$ , and try to prove it for  $k + 1$ . Since  $k < i$  and  $k < r$ , both  $x_k$  and  $y_k$  must bump elements out of row  $k$  into row  $k + 1$ . The element  $y_{k+1}$  bumped by  $y_k$  occurs in column  $s_k$ , which is weakly left of the cell  $(k, j_k)$  occupied by  $x_k$  in  $T \leftarrow x$ . Therefore,  $y_{k+1} \leq x_k$ , which is in turn *strictly* less than  $x_{k+1}$ , the original occupant of cell  $(k, j_k)$  in  $T$ . So  $x_{k+1} > y_{k+1}$ . Repeating the argument used in the first row for row  $k + 1$ , we now see that  $y_{k+1}$  must bump something in row  $k + 1$  into row  $k + 2$  (so that  $r > k + 1$ ), and  $s_{k+1} \leq j_{k+1}$ . This completes the induction. Taking  $k = i$ , we finally conclude that  $r > i$  and  $j = j_i \geq s_i \geq s_r = s$ .  $\square$

## 10.12 Pieri Rules

Iteration of the bumping comparison theorem proves the following result.

**10.63. Theorem: Inserting a Monotone Sequence into a Tableau.** Let  $T$  be a semistandard tableau of shape  $\mu$ , and let  $S$  be the semistandard tableau obtained from  $T$  by insertion of  $z_1, z_2, \dots, z_k$  (in this order); we write  $S = (T \leftarrow z_1 z_2 \cdots z_k)$  in this situation. Let  $\nu$  be the shape of  $S$ .

- (a) If  $z_1 \leq z_2 \leq \cdots \leq z_k$ , then  $\nu/\mu$  is a horizontal strip of size  $k$ .
- (b) If  $z_1 > z_2 > \cdots > z_k$ , then  $\nu/\mu$  is a vertical strip of size  $k$ .

Since tableau insertion is reversible given the location of the new box, we can also reverse the insertion of a monotone sequence, in the following sense.

**10.64. Theorem: Reverse Insertion of a Monotone Sequence.** Suppose  $\mu$  and  $\nu$  are given partitions, and  $S$  is any semistandard tableau of shape  $\nu$ .

- (a) If  $\nu/\mu$  is a horizontal strip of size  $k$ , then there exists a unique sequence  $z_1 \leq z_2 \leq \cdots \leq$



$z_k$  and a unique semistandard tableau  $T$  of shape  $\mu$  such that  $S = (T \leftarrow z_1 z_2 \cdots z_k)$ .

(b) If  $\nu/\mu$  is a vertical strip of size  $k$ , then there exists a unique sequence  $z_1 > z_2 > \cdots > z_k$  and a unique semistandard tableau  $T$  of shape  $\mu$  such that  $S = (T \leftarrow z_1 z_2 \cdots z_k)$ .

*Proof.* To prove the existence of  $T$  and the  $z_i$ 's in part (a), we repeatedly perform reverse tableau insertion, erasing each cell in the horizontal strip  $\nu/\mu$  from right to left. This produces a sequence of elements  $z_k, \dots, z_2, z_1$  and a semistandard tableau  $T$  of shape  $\mu$  such that  $(T \leftarrow z_1 z_2 \cdots z_k) = S$ . Keeping in mind the relative locations of the new boxes created by  $z_i$  and  $z_{i+1}$ , we see from the bumping comparison theorem that  $z_i \leq z_{i+1}$  for all  $i$ .

As for uniqueness, suppose  $T'$  and  $z'_1 \leq z'_2 \leq \cdots \leq z'_k$  also satisfy  $S = (T' \leftarrow z'_1 z'_2 \cdots z'_k)$ . Since  $z'_1 \leq z'_2 \leq \cdots \leq z'_k$ , the bumping comparison theorem shows that the insertion of the  $z'_i$ 's creates the new boxes of  $\nu/\mu$  in order from left to right, just as the insertion of the  $z_i$ 's does. Write  $T_0 = T$ ,  $T_i = (T \leftarrow z_1 z_2 \cdots z_i)$ ,  $T'_0 = T'$ , and  $T'_i = (T' \leftarrow z'_1 z'_2 \cdots z'_i)$ . Since reverse tableau insertion produces a unique answer given the location of the new box, one now sees by reverse induction on  $i$  that  $T_i = T'_i$  and  $z_i = z'_i$  for  $k \geq i \geq 0$ .

Part (b) is proved similarly.  $\square$

**10.65. Theorem: Pieri Rules.** Given an integer partition  $\mu$  and positive integer  $k$ , let  $H_k(\mu)$  consist of all partitions  $\nu$  such that  $\nu/\mu$  is a horizontal strip of size  $k$ , and let  $V_k(\mu)$  consist of all partitions  $\nu$  such that  $\nu/\mu$  is a vertical strip of size  $k$ . For every ordered set  $X$ , there are weight-preserving bijections

$$F : \text{SSYT}_X(\mu) \times \text{SSYT}_X((k)) \rightarrow \bigcup_{\nu \in H_k(\mu)} \text{SSYT}_X(\nu);$$

$$G : \text{SSYT}_X(\mu) \times \text{SSYT}_X((1^k)) \rightarrow \bigcup_{\nu \in V_k(\mu)} \text{SSYT}_X(\nu).$$

Consequently, we have the *Pieri rules* in  $\Lambda_N$ :

$$s_\mu h_k = \sum_{\nu \in H_k(\mu)} s_\nu; \quad s_\mu e_k = \sum_{\nu \in V_k(\mu)} s_\nu.$$

*Proof.* Recall that a semistandard tableau of shape  $(k)$  can be identified with a weakly increasing sequence  $z_1 \leq z_2 \leq \cdots \leq z_k$  of elements of  $X$ . So, we can define  $F(T, z_1 z_2 \cdots z_k) = (T \leftarrow z_1 z_2 \cdots z_k)$ . By 10.63,  $F$  does map into the stated codomain. Then 10.64 shows that  $F$  is a bijection. Moreover,  $F$  is weight-preserving, since the content monomial of  $(T \leftarrow z_1 z_2 \cdots z_k)$  is  $x^{c(T)} x_{z_1} \cdots x_{z_k}$ .

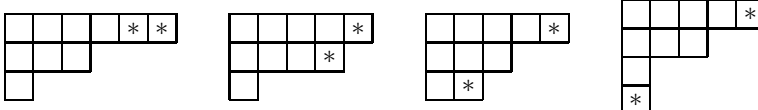
Similarly, a semistandard tableau of shape  $(1^k)$  can be identified with a strictly increasing sequence  $y_1 < y_2 < \cdots < y_k$ . Reversing this gives a strictly decreasing sequence. So we define  $G(T, y_1 y_2 \cdots y_k) = (T \leftarrow y_k \cdots y_2 y_1)$ . As above, 10.63 and 10.64 show that  $G$  is a well-defined bijection.

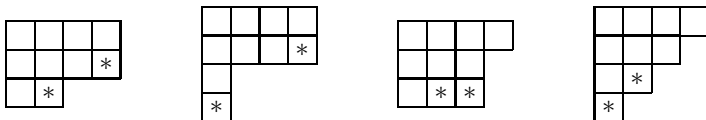
Finally, the Pieri rules follow by passing from weighted sets to generating functions, keeping in mind the sum and product rules for weighted sets, and using  $h_k = s_{(k)}$  and  $e_k = s_{(1^k)}$ .  $\square$

**10.66. Example.** We have

$$s_{(4,3,1)} h_2 = s_{(6,3,1)} + s_{(5,4,1)} + s_{(5,3,2)} + s_{(5,3,1,1)} + s_{(4,4,2)} + s_{(4,4,1,1)} + s_{(4,3,3)} + s_{(4,3,2,1)},$$

as we see by drawing the following pictures:

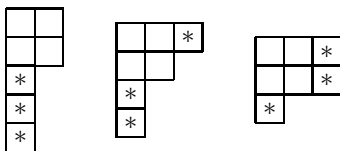




Similarly, we find that

$$s_{(2,2)}e_3 = s_{(2,2,1,1,1)} + s_{(3,2,1,1)} + s_{(3,3,1)}$$

by adding vertical strips to  $\text{dg}((2,2))$  as shown here:



### 10.13 Schur Expansion of $h_\alpha$

Iteration of the Pieri rules lets us compute the Schur expansions of products of the form  $s_\mu h_{\alpha_1} h_{\alpha_2} \cdots h_{\alpha_s}$ , or  $s_\mu e_{\alpha_1} e_{\alpha_2} \cdots e_{\alpha_s}$ , or even “mixed products” involving both  $h$ ’s and  $e$ ’s. Taking  $\mu = 0$ , so that  $s_\mu = 1$ , we obtain in particular the expansions of  $h_\alpha$  and  $e_\alpha$  into sums of Schur polynomials. As we will see, examination of these expansions will lead to another occurrence of the Kostka matrix (cf. 10.50).

**10.67. Example.** Let us use the Pieri rule to find the Schur expansion of  $h_{(2,1,3)} = h_2 h_1 h_3$ . To start, recall that  $h_2 = s_{(2)}$ . Adding one box to  $\text{dg}((2))$  in all possible ways gives

$$h_2 h_1 = s_{(3)} + s_{(2,1)}.$$

Now we add a horizontal strip of size 3 in all possible ways to get

$$\begin{aligned} h_2 h_1 h_3 &= s_{(3)} h_3 + s_{(2,1)} h_3 \\ &= [s_{(6)} + s_{(5,1)} + s_{(4,2)} + s_{(3,3)}] + [s_{(5,1)} + s_{(4,2)} + s_{(4,1,1)} + s_{(3,2,1)}] \\ &= s_{(6)} + 2s_{(5,1)} + 2s_{(4,2)} + s_{(4,1,1)} + s_{(3,3)} + s_{(3,2,1)}. \end{aligned}$$

Observe that the Schur polynomials  $s_{(5,1)}$  and  $s_{(4,2)}$  each occurred twice in the final expansion. Now, consider the computation of  $h_{(2,3,1)} = h_2 h_3 h_1$ . Since multiplication of polynomials is commutative, this symmetric polynomial must be the same as  $h_{(2,1,3)}$ . But the computations with the Pieri rule involve different intermediate objects. We initially calculate

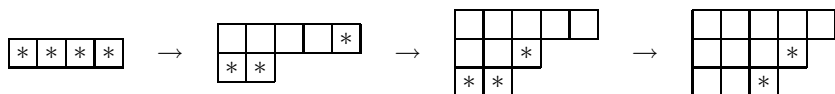
$$h_2 h_3 = s_{(2)} h_3 = s_{(5)} + s_{(4,1)} + s_{(3,2)}.$$

Continuing by multiplying by  $h_1$  gives

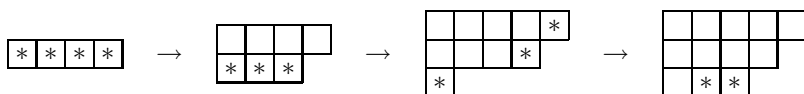
$$\begin{aligned} h_2 h_3 h_1 &= s_{(5)} h_1 + s_{(4,1)} h_1 + s_{(3,2)} h_1 \\ &= [s_{(6)} + s_{(5,1)}] + [s_{(5,1)} + s_{(4,2)} + s_{(4,1,1)}] + [s_{(4,2)} + s_{(3,3)} + s_{(3,2,1)}], \end{aligned}$$

which is the same as the previous answer after collecting terms. As an exercise, the reader is invited to compute  $h_{(3,2,1)} = h_3 h_2 h_1$  and verify that the final answer is again the same.

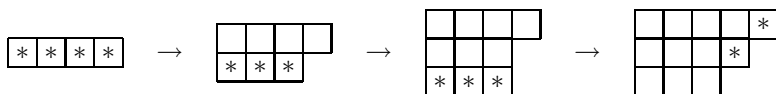
**10.68. Example.** We have seen that a given Schur polynomial may appear several times in the Schur expansion of  $h_\alpha$ . Is there some way to find the coefficient of a particular Schur polynomial in this expansion, without writing down all the shapes generated by iteration of the Pieri rule? To answer this question, consider the problem of finding the coefficient of  $s_{(5,4,3)}$  when  $h_{(4,3,3,2)}$  is expanded into a sum of Schur polynomials. Consider the shapes that appear when we repeatedly use the Pieri rule on the product  $h_4 h_3 h_3 h_2$ . Initially, we have a single shape (4) corresponding to  $h_4$ . Next, we add a horizontal strip of size 3 in all possible ways. Then we add another horizontal strip of size 3 in all possible ways. Finally, we add a horizontal strip of size 2 in all possible ways. The coefficient we seek is *the number of ways that the shape (5, 4, 3) can be built by making the ordered sequence of choices just described*. For example, here is one choice sequence that leads to the shape (5, 4, 3):



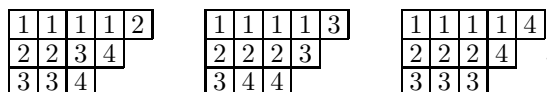
Here is a second choice sequence that leads to the same shape:



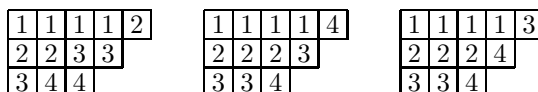
Here is a third choice sequence that leads to the same shape:



Now comes the key observation. We have exhibited each choice sequence by drawing a succession of shapes showing the sequential addition of each horizontal strip. The same information can be encoded by drawing *one* copy of the final shape (5, 4, 3) and putting a label in each box to show which horizontal strip caused that box to first appear in the shape. For example, the three choice sequences displayed above are encoded (in order) by the following three objects:



We have just drawn three semistandard tableaux of shape (5, 4, 3) and content (4, 3, 3, 2)! By definition of the encoding just described, we see that every choice sequence under consideration will be encoded by some tableau of content (4, 3, 3, 2). Since we build the tableau by adding horizontal strips one at a time using increasing labels, it follows that the tableau we get will always be *semistandard*. Finally, we can go backwards in the sense that *any* semistandard tableau of content (4, 3, 3, 2) can be built uniquely by choosing a succession of horizontal strips that tells us where the 1's, 2's, 3's and 4's appear in the tableau. To summarize these remarks, our encoding scheme proves that *the coefficient of  $s_{(5,4,3)}$  in the Schur expansion of  $h_{(4,3,3,2)}$  is the number of semistandard tableaux of shape (5, 4, 3) and content (4, 3, 3, 2)*. In addition to the three semistandard tableaux already drawn, we have the following tableaux of this shape and content:



So the desired coefficient in this example is six.

The argument in the last example generalizes to prove the following result.

**10.69. Theorem: Schur Expansion of  $h_\alpha$ .** Let  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s)$  be any sequence of nonnegative integers with sum  $k$ . Then

$$h_\alpha(x_1, \dots, x_N) = \sum_{\lambda \in \text{Par}_N(k)} K_{\lambda, \alpha} s_\lambda(x_1, \dots, x_N).$$

(It is also permissible to sum over  $\text{Par}(k)$  here.)

*Proof.* By the Pieri rule, the coefficient of  $s_\lambda$  in  $h_\alpha$  is the number of sequences of partitions

$$0 = \mu^0 \subseteq \mu^1 \subseteq \mu^2 \subseteq \dots \subseteq \mu^s = \lambda \quad (10.6)$$

such that the skew shape  $\mu^i / \mu^{i-1}$  is a horizontal strip of size  $\alpha_i$ , for  $1 \leq i \leq s$ . (This is a formal way of describing which horizontal strips we choose at each application of the Pieri rule to the product  $h_\alpha$ .) On the other hand,  $K_{\lambda, \alpha}$  is the number of semistandard tableaux of shape  $\lambda$  and content  $\alpha$ . There is a bijection between the sequences (10.6) and these tableaux, defined by filling each strip  $\mu^i / \mu^{i-1}$  with  $\alpha_i$  copies of the letter  $i$ . The resulting tableau has content  $\alpha$  and is semistandard. The inverse map sends a semistandard tableau  $T$  to the sequence  $(\mu^i : 0 \leq i \leq s)$ , where  $\mu^i$  consists of the cells of  $T$  containing symbols in  $\{1, 2, \dots, i\}$ .  $\square$

**10.70. Remark.** Suppose  $\alpha, \beta$  are sequences such that  $\text{sort}(\alpha) = \text{sort}(\beta)$ . Note that  $h_\alpha = h_\beta$  since multiplication of polynomials is commutative. Expanding each side into Schur polynomials gives

$$\sum_{\lambda \vdash k} K_{\lambda, \alpha} s_\lambda(x_1, \dots, x_N) = \sum_{\lambda \vdash k} K_{\lambda, \beta} s_\lambda(x_1, \dots, x_N).$$

For  $N \geq k$ , the Schur polynomials appearing here will be linearly independent by 10.49. So  $K_{\lambda, \alpha} = K_{\lambda, \beta}$  for all  $\lambda$ , in confirmation of 10.33. (This remark leads to an algebraic proof of 10.33, provided one first gives an algebraic proof of the linear independence of Schur polynomials.)

**10.71. Remark.** The previous theorem and remark extend to skew shapes as follows. First,

$$s_\mu h_\alpha = \sum_{\lambda \in \text{Par}_N} K_{\lambda/\mu, \alpha} s_\lambda.$$

One need only change  $\mu^0$  from 0 to  $\mu$  in the proof above. Second, if  $\text{sort}(\alpha) = \text{sort}(\beta)$ , then  $K_{\lambda/\mu, \alpha} = K_{\lambda/\mu, \beta}$ .

**10.72. Theorem: Complete Homogeneous Basis of  $\Lambda_N^k$ .** For all  $k, N \in \mathbb{N}$ , the set of complete homogeneous polynomials

$$\{h_\mu(x_1, \dots, x_N) : \mu \in \text{Par}_N(k)\} \subseteq K[x_1, \dots, x_N]$$

is a basis of the  $K$ -vector space  $\Lambda_N^k$ .

*Proof.* Consider column vectors  $\mathbf{S} = (s_\lambda(x_1, \dots, x_N) : \lambda \in \text{Par}_N(k))$  and  $\mathbf{H} = (h_\mu(x_1, \dots, x_N) : \mu \in \text{Par}_N(k))$ , where the entries are listed in lexicographic order. As in the proof of 10.49, let  $\mathbf{K} = (K_{\lambda, \mu})$  be the Kostka matrix with rows and columns indexed by partitions in  $\text{Par}_N(k)$  in lexicographic order. Recall from 10.47 that  $\mathbf{K}$  is a lower-triangular matrix with 1's on the main diagonal. In matrix form, 10.69 asserts that  $\mathbf{H} = \mathbf{K}^t \mathbf{S}$ , where  $\mathbf{K}^t$  is the transpose of the Kostka matrix. This transpose is upper-triangular with 1's on the main diagonal, hence is invertible. Since  $\mathbf{H}$  is obtained from  $\mathbf{S}$  by application of an invertible matrix of scalars, we see that the elements of  $\mathbf{H}$  form a basis by the same reasoning used in the proof of 10.49 (cf. 10.178).  $\square$

**10.73. Remark.** Combining 10.72 with 10.49, we can write  $\mathbf{H} = (\mathbf{K}^t \mathbf{K}) \mathbf{M}$ , where  $\mathbf{M}$  is the vector of monomial symmetric polynomials indexed by  $\text{Par}_N(k)$ . This matrix equation gives the monomial expansion of the complete homogeneous symmetric polynomials  $h_\mu$ .

## 10.14 Schur Expansion of $e_\alpha$

Now we turn to the elementary symmetric polynomials  $e_\alpha$ . We can iterate the Pieri rule as we did for  $h_\alpha$ , but here we must add vertical strips at each stage.

**10.74. Example.** Let us compute the Schur expansion of  $e_{(2,2,2)} = e_2 e_2 e_2$ . First,  $e_2 e_2 = s_{(1,1)} e_2 = s_{(2,2)} + s_{(2,1,1)} + s_{(1,1,1,1)}$ . Next,

$$\begin{aligned} e_2 e_2 e_2 &= [s_{(3,3)} + s_{(3,2,1)} + s_{(2,2,1,1)}] \\ &\quad + [s_{(3,2,1)} + s_{(3,1,1,1)} + s_{(2,2,2)} + s_{(2,2,1,1)} + s_{(2,1,1,1,1)}] \\ &\quad + [s_{(2,2,1,1)} + s_{(2,1,1,1,1)} + s_{(1,1,1,1,1,1)}] \\ &= s_{(3,3)} + 2s_{(3,2,1)} + s_{(3,1,1,1)} + s_{(2,2,2)} + 3s_{(2,2,1,1)} + 2s_{(2,1^4)} + s_{(1^6)}. \end{aligned}$$

As in the case of  $h_\alpha$ , we can use tableaux to encode the sequence of vertical strips chosen in the repeated application of the Pieri rules. For example, the following tableaux encode the three choice sequences that lead to the shape  $(2, 2, 1, 1)$  in the expansion of  $e_{(2,2,2)}$ :

1	2	1	2	1	3
1	2	1	3	1	3
3		2		2	
3		3		2	

Evidently, these tableaux are not semistandard (column-strict). However, transposing the diagrams will produce semistandard tableaux of shape  $(2, 2, 1, 1)' = (4, 2)$  and content  $(2, 2, 2)$ , as shown here:

1	1	3	3	1	1	2	3	1	1	2	2
2	2			2	3			3	3		

This encoding gives a bijection from the relevant choice sequences to the collection of semistandard tableaux of this shape and content. So the coefficient of  $s_{(2,2,1,1)}$  in the Schur expansion of  $e_{(2,2,2)}$  is the Kostka number  $K_{(4,2),(2,2,2)} = 3$ . This argument generalizes to prove the following theorem.

**10.75. Theorem: Schur Expansion of  $e_\alpha$ .** Let  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s)$  be any sequence of nonnegative integers with sum  $k$ . Then

$$e_\alpha(x_1, \dots, x_N) = \sum_{\lambda \in \text{Par}_N(k)} K_{\lambda', \alpha} s_\lambda(x_1, \dots, x_N) = \sum_{\nu \in \text{Par}_N(k)'} K_{\nu, \alpha} s_{\nu'}(x_1, \dots, x_N).$$

**10.76. Remark.** We have written  $\text{Par}_N(k)'$  for the set  $\{\lambda' : \lambda \in \text{Par}_N(k)\}$ . Since conjugation of a partition interchanges the number of parts with the length of the largest part, we have

$$\text{Par}_N(k)' = \{\nu \in \text{Par}(k) : \nu_1 \leq N\} = \{\nu \in \text{Par}(k) : \nu_i \leq N \text{ for all } i \geq 1\}.$$

It is also permissible to sum over all partitions of  $k$  in the theorem, since this will only add zero terms to the sum. If the number of variables is large enough ( $N \geq k$ ), then we will already be summing over all partitions of  $k$ .

**10.77. Theorem: Elementary Basis of  $\Lambda_N^k$ .** For all  $k, N \in \mathbb{N}$ , the set of elementary symmetric polynomials

$$\{e_\mu(x_1, \dots, x_N) : \mu \in \text{Par}_N(k)'\} = \{e_{\mu'}(x_1, \dots, x_N) : \mu \in \text{Par}_N(k)\} \subseteq K[x_1, \dots, x_N]$$

is a basis of the  $K$ -vector space  $\Lambda_N^k$ . Consequently, the set of all polynomials  $e_1^{i_1} \cdots e_N^{i_N}$ , where the  $i_j$  are arbitrary nonnegative integers, is a basis of  $\Lambda_N$ .

*Proof.* We use the same matrix argument employed earlier, suitably adjusted to account for the shape conjugation. As in the past, let us index the rows and columns of matrices and vectors by the partitions in  $\text{Par}_N(k)$ , listed in lexicographic order. Introduce column vectors  $\mathbf{S} = (s_\lambda(x_1, \dots, x_N) : \lambda \in \text{Par}_N(k))$  and  $\mathbf{E} = (e_{\mu'}(x_1, \dots, x_N) : \mu \in \text{Par}_N(k))$ . Next, consider the modified Kostka matrix  $\hat{\mathbf{K}}$  whose entry in row  $\mu$  and column  $\lambda$  is  $K_{\lambda', \mu'}$ . Now 10.75 asserts that

$$e_{\mu'}(x_1, \dots, x_N) = \sum_{\lambda \in \text{Par}_N(k)} K_{\lambda', \mu'} s_\lambda(x_1, \dots, x_N).$$

By definition of matrix-vector multiplication, the equations just written are equivalent to  $\mathbf{E} = \hat{\mathbf{K}}\mathbf{S}$ . Since the entries of  $\mathbf{S}$  are known to be a basis, it suffices (as in the proofs of 10.49 and 10.72) to argue that  $\hat{\mathbf{K}}$  is a triangular matrix with 1's on the diagonal. There are 1's on the diagonal, since  $K_{\mu', \mu'} = 1$ . On the other hand, by 10.45, we have the implications

$$\hat{\mathbf{K}}(\mu, \lambda) \neq 0 \Rightarrow K_{\lambda', \mu'} \neq 0 \Rightarrow \mu' \trianglelefteq \lambda' \Rightarrow \lambda \leq \mu \Rightarrow \lambda \leq_{\text{lex}} \mu.$$

So  $\hat{\mathbf{K}}$  is lower-triangular. The final statement about the basis of  $\Lambda_N$  follows by writing partitions in  $\text{Par}'_N$  in the form  $1^{i_1} 2^{i_2} \cdots N^{i_N}$  and noting that the vector space  $\Lambda_N$  is the direct sum of the subspaces  $\Lambda_N^k$ .  $\square$

**10.78. Remark.** Combining this theorem with 10.49, we can write  $\mathbf{E} = (\hat{\mathbf{K}}\mathbf{K})\mathbf{M}$ , where  $\mathbf{M}$  is the vector of monomial symmetric polynomials indexed by  $\text{Par}_N(k)$ . This matrix equation gives the monomial expansion of elementary symmetric polynomials.

## 10.15 Algebraic Independence

We will use 10.77 to obtain structural information about the ring  $\Lambda_N$  of symmetric polynomials in  $N$  variables. First we need the following definition.

**10.79. Definition: Algebraic Independence.** Let  $A$  be a commutative ring containing  $K$ , and let  $(z_1, \dots, z_N)$  be a list of elements of  $A$ . We say that  $(z_1, \dots, z_N)$  is *algebraically independent over  $K$*  iff the collection of monomials

$$\{z^\alpha = z_1^{\alpha_1} z_2^{\alpha_2} \cdots z_N^{\alpha_N} : \alpha \in \mathbb{N}^N\}$$

is linearly independent over  $K$ . This means that whenever a finite  $K$ -linear combination of the  $z^\alpha$ 's is zero, say

$$\sum_{\alpha} c_{\alpha} z^{\alpha} = 0 \quad (c_{\alpha} \in K),$$

then all  $c_{\alpha}$ 's must be zero.

Here is yet another formulation of the definition. Let  $K[Z_1, \dots, Z_N]$  be a polynomial ring in  $N$  indeterminates. Given any list  $(z_1, \dots, z_N) \in A^N$ , we get an evaluation homomorphism  $T : K[Z_1, \dots, Z_N] \rightarrow A$  that sends each  $Z_i$  to  $z_i$  (see 7.102). One can check that the image of  $T$  is the subring  $B$  of  $A$  generated by  $K$  and the  $z_i$ 's. On the other hand, the kernel of  $T$  consists precisely of the polynomials  $\sum_{\alpha} c_{\alpha} Z^{\alpha}$  such that  $\sum_{\alpha} c_{\alpha} z^{\alpha} = 0$ . So, the  $z_i$ 's are algebraically independent over  $K$  iff  $\ker(T) = \{0\}$  iff  $T$  is *injective*. In this case,  $T$  (with codomain restricted to  $B$ ) is a ring isomorphism, and  $K[Z_1, \dots, Z_N] \cong B$ . So we may identify  $Z_i$  with  $z_i$  and write  $B = K[z_1, \dots, z_N]$ .

**10.80. Example.** Let  $K[x_1, \dots, x_N]$  be a polynomial ring in indeterminates  $x_i$ . By the very definition of polynomials,  $\sum_{\alpha} c_{\alpha} x^{\alpha} = 0$  implies all  $c_{\alpha}$  are zero. So the indeterminates  $x_1, \dots, x_N$  are algebraically independent over  $K$ . The evaluation map  $T$  above can be taken to be the identity function on  $K[x_1, \dots, x_N]$ . On the other hand, consider the three polynomials  $z_1 = x_1 + x_2$ ,  $z_2 = x_1^2 + x_2^2$ , and  $z_3 = x_1^3 + x_2^3$ . The elements  $z_1, z_2, z_3$  are *linearly* independent over  $K$ , as one may check. However, they are not *algebraically* independent over  $K$ , because of the relation

$$1z_1^3 - 3z_1z_2 + 2z_3 = 0.$$

Later, we will see that  $z_1$  and  $z_2$  are algebraically independent over  $K$ .

By 10.77,  $\Lambda_N$  is the subring of  $K[x_1, \dots, x_N]$  generated by  $K$  and the elementary symmetric polynomials. Combining the last part of 10.77 with 10.79, we deduce the following structural result.

**10.81. Fundamental Theorem of Symmetric Polynomials.** The elementary symmetric polynomials

$$\{e_i(x_1, \dots, x_N) : 1 \leq i \leq N\} \subseteq K[x_1, \dots, x_N]$$

are algebraically independent over  $K$ . Furthermore, if  $K[E_1, \dots, E_N]$  is another polynomial ring, then the evaluation map  $T : K[E_1, \dots, E_N] \rightarrow \Lambda_N$  sending  $E_i$  to  $e_i(x_1, \dots, x_N)$  is an isomorphism of rings and  $K$ -vector spaces. So, for every symmetric polynomial  $f(x_1, \dots, x_N)$ , there exists a unique polynomial  $g(E_1, \dots, E_N)$  such that  $f = T(g) = g(e_1, \dots, e_N)$ .

**10.82. Remark.** An algorithmic proof of the existence assertion in the fundamental theorem is sketched in 10.211.

## 10.16 Power-Sum Symmetric Polynomials

Recall that the *power-sum* symmetric polynomials in  $N$  variables are defined by setting  $p_k(x_1, \dots, x_N) = \sum_{i=1}^N x_i^k$  for all  $k \geq 1$  and  $p_{\alpha}(x_1, \dots, x_N) = \prod_{j \geq 1} p_{\alpha_j}(x_1, \dots, x_N)$ . It turns out that the polynomials  $(p_1, \dots, p_N)$  are algebraically independent over  $K$ . One way to prove this is to invoke the following determinant criterion for algebraic independence.

**10.83. Theorem: Determinant Test for Algebraic Independence.** Let  $g_1, \dots, g_N$  be  $N$  polynomials in  $K[x_1, \dots, x_N]$ . Let  $\mathbf{A}$  be the  $N \times N$  matrix whose  $j, k$ -entry is the formal partial derivative  $D_j g_k = \partial g_k / \partial x_j$  (see 7.103), and let  $J \in K[x_1, \dots, x_N]$  be the determinant of  $\mathbf{A}$  (see 9.37). If  $J \neq 0$ , then  $g_1, \dots, g_N$  are algebraically independent over  $K$ .

*Proof.* We prove the contrapositive. Assume  $g_1, \dots, g_N$  are algebraically dependent over  $K$ . Then there exist nonzero polynomials  $h \in K[Z_1, \dots, Z_N]$  such that  $h(g_1, \dots, g_N) = 0$ . Choose such an  $h$  whose total degree in the  $Z_i$ 's is as small as possible. We can take the partial derivative of  $h(g_1, \dots, g_N)$  with respect to  $x_j$  by applying the formal multivariable chain rule (see 7.104), obtaining the relations

$$\sum_{k=1}^N (D_k h)(g_1, \dots, g_N) \frac{\partial g_k}{\partial x_j} = 0 \quad (1 \leq j \leq N).$$

Let  $\mathbf{v}$  be the column vector whose  $k$ th entry is  $(D_k h)(g_1, \dots, g_N)$ . The preceding relations are equivalent to the matrix identity  $\mathbf{A}\mathbf{v} = \mathbf{0}$ . We will show that  $\mathbf{v}$  is not the zero vector. Since  $h \in K[Z_1, \dots, Z_N]$  is nonzero and  $K$  contains  $\mathbb{Q}$ , at least one partial derivative  $D_k h \in K[Z_1, \dots, Z_N]$  must be nonzero. Given such a  $k$  with  $D_k h \neq 0$ , the total degree of  $D_k h$  in the  $Z_i$ 's must be lower than the total degree of  $h$ . By choice of  $h$ , it follows that  $(D_k h)(g_1, \dots, g_N)$  is nonzero in  $K[x_1, \dots, x_N]$ . This polynomial is the  $k$ th entry of  $\mathbf{v}$ , so  $\mathbf{v} \neq \mathbf{0}$ . Now  $\mathbf{A}\mathbf{v} = \mathbf{0}$  forces  $\mathbf{A}$  to be a singular matrix, so  $J = \det(\mathbf{A}) = 0$  by a theorem of linear algebra.  $\square$

**10.84. Remark.** The converse of 10.83 is also true: if  $g_1, \dots, g_N$  are algebraically independent, then the “Jacobian”  $J$  will be nonzero. This fact is not needed in the sequel, so we omit the proof.

**10.85. Theorem: Algebraic Independence of Power-Sums.** Let  $K$  be a field containing  $\mathbb{Q}$ . The power-sum polynomials

$$\{p_k(x_1, \dots, x_N) : 1 \leq k \leq N\} \subseteq K[x_1, \dots, x_N]$$

are algebraically independent over  $K$ .

*Proof.* We use the determinant criterion in 10.83. The  $j, k$ -entry of the matrix  $\mathbf{A}$  is

$$D_j p_k = \frac{\partial}{\partial x_j} (x_1^k + x_2^k + \dots + x_j^k + \dots + x_N^k) = kx_j^{k-1}.$$

Accordingly,  $J = \det \|kx_j^{k-1}\|_{1 \leq j, k \leq N}$ . For each column  $k$ , we may factor out the scalar  $k$  to see that  $J = N! \det \|x_j^{k-1}\|$ . The resulting determinant is called a Vandermonde determinant. This determinant evaluates to  $\pm \prod_{1 \leq r < s \leq N} (x_r - x_s)$ , which is a nonzero polynomial (see §12.9 for a combinatorial proof of this formula). Since  $K$  contains  $\mathbb{Q}$ , the scalar  $N!$  is not zero in  $K$ . We conclude that  $J \neq 0$ , which proves the result.  $\square$

Now that we know that the  $p_k$ 's are algebraically independent, we can obtain power-sum bases for the vector spaces  $\Lambda_N^k$  and  $\Lambda_N$ .

**10.86. Theorem: Power-Sum Basis.** Let  $K$  be a field containing  $\mathbb{Q}$ . For all  $k, N \in \mathbb{N}$ , the collection

$$\{p_\mu(x_1, \dots, x_N) : \mu \in \text{Par}_N(k)'\} \subseteq K[x_1, \dots, x_N]$$

is a basis of the  $K$ -vector space  $\Lambda_N^k$ . The collection  $\{p_1^{i_1} \cdots p_N^{i_N} : i_j \geq 0\}$  is a basis of the  $K$ -vector space  $\Lambda_N$ . Letting  $P_1, \dots, P_N$  be new indeterminates, there is an isomorphism of rings and  $K$ -vector spaces  $T : K[P_1, \dots, P_N] \rightarrow \Lambda_N$  such that  $T(P_i) = p_i(x_1, \dots, x_N)$ . So, for every symmetric polynomial  $f(x_1, \dots, x_N)$ , there exists a unique polynomial  $g$  with  $f = g(p_1, \dots, p_N)$ .



### 10.17 Relations between $e$ 's and $h$ 's

We have seen that, in the polynomial ring  $K[x_1, \dots, x_N]$ , the lists  $(e_1, \dots, e_N)$  and  $(p_1, \dots, p_N)$  are each algebraically independent. One might wonder if the polynomials  $(h_1, \dots, h_N)$  are also algebraically independent over  $K$ . This would follow (as it did for the  $e$ 's) if we knew that  $\{h_\mu : \mu \in \text{Par}_N(k)'\}$  was a basis of  $\Lambda_N^k$  for all  $k$ . However, the basis we found in 10.72 was  $\{h_\mu : \mu \in \text{Par}_N(k)\}$ , which is indexed by partitions of  $k$  with at most  $N$  parts, instead of partitions of  $k$  with each part at most  $N$ . The next result will allow us to overcome this difficulty by providing equations relating  $e_1, \dots, e_N$  to  $h_1, \dots, h_N$ .

**10.87. Theorem: Recursion for  $e_i$ 's and  $h_j$ 's.** For all  $m, N \in \mathbb{N}$ , we have the identity

$$\sum_{i=0}^m (-1)^i e_i(x_1, \dots, x_N) h_{m-i}(x_1, \dots, x_N) = \chi(m=0). \quad (10.7)$$

*Proof.* If  $m = 0$ , the identity reads  $1 = 1$ , so let us assume  $m > 0$ . We can model the left side of the identity using a collection  $Z$  of signed, weighted objects. A typical object in  $Z$  is a triple  $z = (i, S, T)$ , where  $0 \leq i \leq m$ ,  $S \in \text{SSYT}_N((1^i))$ , and  $T \in \text{SSYT}_N((m-i))$ . The *weight* of  $(i, S, T)$  is  $x^{c(S)}x^{c(T)}$ , and the *sign* of  $(i, S, T)$  is  $(-1)^i$ . For example, taking  $N = 9$  and  $m = 7$ , a typical object in  $Z$  is

$$z = \left( 3, \begin{array}{|c|} \hline 2 \\ \hline 4 \\ \hline 7 \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline 3 & 3 & 4 & 6 \\ \hline \end{array} \right).$$

The signed weight of this object is  $(-1)^3(x_2x_4x_7)(x_3^2x_4x_6) = -x_2x_3^2x_4^2x_6x_7$ . Recalling that  $e_i = s_{(1^i)}$  and  $h_{m-i} = s_{(m-i)}$ , we see that the left side of (10.7) is precisely

$$\sum_{z \in Z} \text{sgn}(z) \text{wt}(z).$$

To prove this expression is zero, we define a sign-reversing involution  $I : Z \rightarrow Z$  with no fixed points. Given  $z = (i, S, T) \in Z$ , we compute  $I(z)$  as follows. Let  $j = S((1, 1))$  be the smallest entry in  $S$ , and let  $k = T((1, 1))$  be the leftmost entry in  $T$ . If  $i = 0$ , then  $S$  is empty and  $j$  is undefined; if  $i = m$ , then  $T$  is empty and  $k$  is undefined. Since  $m > 0$ , at least one of  $j$  or  $k$  is defined. If  $j \leq k$  or  $k$  is not defined, move the box containing  $j$  from  $S$  to  $T$ , so that this box is the new leftmost entry in  $T$ , and decrement  $i$  by 1. Otherwise, if  $k < j$  or  $j$  is not defined, move the box containing  $k$  from  $T$  to  $S$ , so that this box is the new topmost box in  $S$ , and increment  $i$  by 1. For example, if  $z$  is the object shown above, then

$$I(z) = \left( 2, \begin{array}{|c|} \hline 4 \\ \hline 7 \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline 2 & 3 & 3 & 4 & 6 \\ \hline \end{array} \right).$$

As another example,

$$I((0, \emptyset, \begin{array}{|c|c|c|c|c|c|c|} \hline 2 & 2 & 3 & 5 & 5 & 7 & 9 \\ \hline \end{array})) = (1, \begin{array}{|c|} \hline 2 \\ \hline \end{array}, \begin{array}{|c|c|c|c|c|c|c|} \hline 2 & 3 & 5 & 5 & 7 & 9 \\ \hline \end{array}).$$

From the definition of  $I$ , we can check that  $I$  does map  $Z$  into  $Z$ , that  $I \circ I = \text{id}_Z$ , that  $I$  is weight-preserving and sign-reversing, and that  $I$  has no fixed points.  $\square$

**10.88. Theorem: Algebraic Independence of  $h$ 's.** For all  $k, N \in \mathbb{N}$ , the collection

$$\{h_\mu(x_1, \dots, x_N) : \mu \in \text{Par}_N(k)'\} \subseteq K[x_1, \dots, x_N]$$

is a basis of the  $K$ -vector space  $\Lambda_N^k$ . Consequently,  $\{h_1^{i_1} \cdots h_N^{i_N} : i_j \geq 0\}$  is a basis of the  $K$ -vector space  $\Lambda_N$ , and  $(h_1, \dots, h_N)$  is algebraically independent over  $K$ . Letting  $H_1, \dots, H_N$  be new indeterminates, there is a ring and  $K$ -vector space isomorphism  $T : K[H_1, \dots, H_N] \rightarrow \Lambda_N$  given by  $T(H_i) = h_i(x_1, \dots, x_N)$ . So, for every symmetric polynomial  $f(x_1, \dots, x_N)$ , there exists a unique polynomial  $g$  with  $f = g(h_1, \dots, h_N)$ .

*Proof.* It suffices to prove the statement about the basis of  $\Lambda_N^k$ , from which the other assertions follow. By 10.72, we know that

$$|\{h_\mu(x_1, \dots, x_N) : \mu \in \text{Par}_N(k)'\}| \leq |\{h_\mu(x_1, \dots, x_N) : \mu \in \text{Par}_N(k)\}| = \dim_K(\Lambda_N^k).$$

So, by a theorem of linear algebra, it is enough to prove that  $\{h_\mu : \mu \in \text{Par}_N(k)'\}$  is a *spanning set* of  $\Lambda_N^k$ . For each  $k \geq 0$ , let  $V_N^k$  be the vector subspace of  $\Lambda_N^k$  spanned by these  $h_\mu$ 's. We want to prove  $V_N^k = \Lambda_N^k$  for all  $k$ . It will suffice to show that  $e_1^{i_1} \cdots e_N^{i_N} \in V_N^k$  for all  $i_1, \dots, i_N$  that sum to  $k$ , since these elementary symmetric polynomials are known to be a basis of  $\Lambda_N^k$ . Now, one can check that  $f \in V_N^k$  and  $g \in V_N^m$  imply  $fg \in V_N^{k+m}$ . (This holds when  $f$  and  $g$  are each products of  $h_1, \dots, h_N$ , and the general case follows by linearity and the distributive law.) Using this remark, we can further reduce to proving that  $e_j(x_1, \dots, x_N) \in V_N^j$  for  $1 \leq j \leq N$ .

We prove this by induction on  $j$ . The result is true for  $j = 1$ , since  $e_1 = \sum_{k=1}^N x_k = h_1 \in V_N^1$ . Assume  $1 < j \leq N$  and the result is known to hold for all smaller values of  $j$ . Taking  $m = j$  in the recursion (10.7), we have

$$e_j = e_{j-1}h_1 - e_{j-2}h_2 + e_{j-3}h_3 - \cdots \pm e_1h_{j-1} \mp h_j.$$

Since  $e_{j-s} \in V_N^{j-s}$  (by induction) and  $h_s \in V_N^s$  (by definition) for  $1 \leq s \leq j$ , each term on the right side lies in  $V_N^j$ . Since  $V_N^j$  is a subspace, it follows that  $e_j \in V_N^j$ , completing the induction.  $\square$

## 10.18 Generating Functions for $e$ 's and $h$ 's

Another approach to the identities (10.7) involves generating functions.

**10.89. Definition:**  $E_N(t)$  and  $H_N(t)$ . For each  $N \geq 1$ , define the polynomial

$$E_N(t) = \prod_{i=1}^N (1 + x_i t) \in F(x_1, \dots, x_N)[t]$$

and the formal power series

$$H_N(t) = \prod_{i=1}^N \frac{1}{1 - x_i t} \in F(x_1, \dots, x_N)[[t]].$$

**10.90. Theorem: Expansion of  $E_N(t)$ .** For all  $N \geq 1$ , we have

$$E_N(t) = \sum_{k=0}^N e_k(x_1, \dots, x_N) t^k.$$

*Proof.* Let us use the generalized distributive law 2.7 to expand the product in the definition of  $E_N(t)$ . We obtain

$$E_N(t) = \prod_{i=1}^N (1 + x_i t) = \sum_{S \subseteq \{1, 2, \dots, N\}} \prod_{i \in S} (x_i t) \prod_{i \notin S} 1.$$

To get terms involving  $t^k$ , we must restrict the sum to subsets  $S$  of size  $k$ . Such subsets can be identified with increasing sequences  $1 \leq i_1 < i_2 < \dots < i_k \leq N$ . Therefore, the coefficient of  $t^k$  in  $E_N(t)$  is

$$\sum_{1 \leq i_1 < i_2 < \dots < i_k \leq N} x_{i_1} x_{i_2} \dots x_{i_k} = e_k(x_1, \dots, x_N). \quad \square$$

**10.91. Theorem: Relation between Roots and Coefficients of a Polynomial.**

Suppose a polynomial  $p = X^N + a_1 X^{N-1} + \dots + a_i X^{N-i} + \dots + a_{N-1} X + a_N \in K[X]$  factors as  $p = (X - r_1)(X - r_2) \dots (X - r_N)$  for some  $r_i \in K$ . For  $1 \leq i \leq N$ ,

$$a_i = (-1)^i e_i(r_1, r_2, \dots, r_N).$$

*Proof.* One can prove this by expanding  $\prod_{i=1}^N (X - r_i)$  with the generalized distributive law. Alternatively, we can deduce the result from 10.90 as follows. Replacing  $t$  by  $1/X$  and  $x_i$  by  $-r_i$  in  $E_N(t)$  gives  $\prod_{i=1}^N (1 - r_i/X) = X^{-N} p$ . Using 10.90, we conclude that

$$p = X^N \sum_{k=0}^n e_k(-r_1, \dots, -r_N) X^{-k} = \sum_{k=0}^N (-1)^k e_k(r_1, \dots, r_N) X^{N-k}.$$

Taking the coefficient of  $X^{N-i}$  gives the result.  $\square$

**10.92. Theorem: Expansion of  $H_N(t)$ .** For all  $N \geq 1$ , we have

$$H_N(t) = \sum_{k=0}^{\infty} h_k(x_1, \dots, x_N) t^k.$$

*Proof.* Using the geometric series formula, we have

$$H_N(t) = \prod_{i=1}^N \frac{1}{1 - x_i t} = \prod_{i=1}^N \sum_{j_i=0}^{\infty} (x_i t)^{j_i}.$$

Next, using the generalized distributive law, we get

$$H_N(t) = \sum_{(j_1, \dots, j_N) \in \mathbb{N}^N} \prod_{i=1}^N x_i^{j_i} t^{j_i} = \sum_{(j_1, \dots, j_N) \in \mathbb{N}^N} t^{j_1 + \dots + j_N} \prod_{i=1}^N x_i^{j_i}.$$

The coefficient of  $t^k$  consists of the sum of all possible monomials in  $x_1, \dots, x_N$  of degree  $k$ , which is precisely  $h_k(x_1, \dots, x_N)$ .  $\square$

Now we can give an algebraic proof of 10.87. Note that

$$H_N(t) E_N(-t) = \prod_{i=1}^N \frac{1}{(1 - x_i t)} \prod_{i=1}^N (1 - x_i t) = 1.$$

Equating the coefficients of  $t^m$  on both sides gives the identities (10.7).

### 10.19 Relations between $p$ 's, $e$ 's, and $h$ 's

In this section, we will study recursions similar to (10.7) that relate the  $h_n$ 's and  $e_n$ 's to the  $p_k$ 's. These recursions can be used to deduce the algebraic independence of the  $p_k$ 's from the algebraic independence of the  $h_n$ 's (or  $e_n$ 's), and vice versa, by adapting the proof of 10.88 to the new recursions.

**10.93. Theorem: Recursion for  $h_i$ 's and  $p_j$ 's.** For all  $n, N \geq 1$ , the following identity is valid in  $\Lambda_N$ :

$$h_0 p_n + h_1 p_{n-1} + h_2 p_{n-2} + \cdots + h_{n-1} p_1 = n h_n. \quad (10.8)$$

*Proof.* Let us interpret each side of the desired equation as the generating function for a suitable collection of weighted objects. For the left side, let  $X$  be the set of all triples  $(k, T, U)$ , where:  $0 \leq k < n$ ;  $T \in \text{SSYT}_N((k))$ ; and  $U$  consists of a row of  $n - k$  boxes all filled with the same integer  $i \leq N$ . The *weight* of such a triple is  $x^{c(T)+c(U)} = x^{c(T)} x_i^{n-k}$ . For example, letting  $n = 8$  and  $N = 9$ , here is a typical object in  $X$  of weight  $x_1^2 x_2 x_3^3 x_4^2$ :

$$z_0 = (5, \boxed{1 \ 1 \ 2 \ 4 \ 4}, \boxed{3 \ 3 \ 3}).$$

For a fixed value of  $k$ , the generating function for the possible  $T$ 's is  $h_k(x_1, \dots, x_N)$  and the generating function for the possible  $U$ 's is  $p_{n-k}(x_1, \dots, x_N)$ . By the sum and product rules for weighted sets, the left side of (10.8) is the generating function for  $X$ .

Now let  $Y$  be the set of all pairs  $(V, j)$ , where  $V \in \text{SSYT}_N((n))$  and  $1 \leq j \leq n$ . We can visualize an object in  $Y$  as a semistandard tableau of shape  $(n)$  in which the  $j$ th cell has been *marked*. For example, here is a typical object in  $Y$  of weight  $x_1^2 x_3^5 x_4$ :

$$y_0 = \boxed{1 \ 1 \ 3 \ 3^* \ 3 \ 3 \ 3 \ 4}.$$

The generating function for the weighted set  $Y$  is  $n h_n(x_1, \dots, x_N)$ .

To prove (10.8), it suffices to define a weight-preserving bijection  $f : X \rightarrow Y$ . Given  $(k, T, U) \in X$ , note that  $U$  consists of a run of  $n - k$  copies of some symbol  $i$ . To compute  $f((k, T, U))$ , mark the first box in  $U$  and splice the boxes of  $U$  into  $T$  in the appropriate position to get a weakly increasing sequence. If  $T$  already contains one or more  $i$ 's, the first box of  $U$  is inserted immediately after these  $i$ 's. For example,

$$f(z_0) = \boxed{1 \ 1 \ 2 \ 3^* \ 3 \ 3 \ 4 \ 4}.$$

This insertion process is reversible, thanks to the marker. More precisely, define  $g : Y \rightarrow X$  as follows. Given  $(V, j) \in Y$ , let  $i$  be the entry in the  $j$ th cell of  $V$ . Starting at cell  $j$  and scanning right, remove each cell equal to  $i$  from  $V$  to get a pair of tableaux  $T$  and  $U$  as in the definition of  $X$ . Define  $g((V, j)) = (k, T, U)$ , where  $k$  is the number of boxes in  $T$ . For example,

$$g(y_0) = (4, \boxed{1 \ 1 \ 3 \ 4}, \boxed{3 \ 3 \ 3 \ 3}).$$

One may check that  $f$  and  $g$  are weight-preserving and are two-sided inverses of each other.  $\square$

**10.94. Theorem: Recursion for  $e_i$ 's and  $p_j$ 's.** For all  $n, N \geq 1$ , the following identity is valid in  $\Lambda_N$ :

$$e_0 p_n - e_1 p_{n-1} + e_2 p_{n-2} - \cdots + (-1)^{n-1} e_{n-1} p_1 = (-1)^{n-1} n e_n. \quad (10.9)$$

*Proof.* This time we interpret each side of the equation using suitable *signed*, weighted objects. For the left side, let  $X$  be the set of all triples  $(k, T, U)$ , where:  $0 \leq k < n$ ;  $T \in \text{SSYT}_N((1^k))$ ; and  $U$  consists of a row of  $n - k$  boxes all filled with the same integer  $i \leq N$ . The *weight* of this triple is  $x^{c(T)+c(U)}$ , and the *sign* of this triple is  $(-1)^k$ . For example, here is a typical object of  $X$  whose signed weight is  $(-1)^4 x_2 x_4^5 x_5 x_7$ :

$$z_0 = \left( 4, \begin{array}{|c|} \hline 2 \\ \hline 4 \\ \hline 5 \\ \hline 7 \\ \hline \end{array}, \begin{array}{|c|c|c|c|c|} \hline 4 & 4 & 4 & 4 & 4 \\ \hline \end{array} \right).$$

Using the sum and product rules for weighted sets, one sees that  $\sum_{z \in X} \text{sgn}(z) \text{wt}(z)$  is the left side of (10.9).

Now let  $Y = \{(T, j) : T \in \text{SSYT}_N((1^n)), 1 \leq j \leq n\}$ . We can think of each element of  $Y$  as a strictly increasing sequence of  $n$  elements of  $\{1, 2, \dots, N\}$  in which one of the elements (the  $j$ th one) has been *marked*. The generating function for the weighted set  $Y$  is  $ne_n(x_1, \dots, x_N)$ .

Let us define a weight-preserving, sign-reversing involution  $I : X \rightarrow X$ . Fix  $(k, T, U) \in X$ . Since  $k < n$ ,  $U$  is not empty; let  $j$  be the integer appearing in each box of  $U$ . The map  $I$  acts as follows. On one hand, if  $k < n - 1$  and  $j$  does not appear in  $T$ , then increase  $k$  by 1, remove one copy of  $j$  from  $U$ , and insert this number in the proper position in  $T$  to get a sorted sequence. On the other hand, if  $j$  does appear in  $T$ , then decrease  $k$  by 1, remove the unique copy of  $j$  from  $T$ , and place another copy of  $j$  in  $U$ . If neither of the two preceding cases occurs,  $(k, T, U)$  is a fixed point of  $I$ . For example,

$$I(z_0) = \left( 3, \begin{array}{|c|} \hline 2 \\ \hline 5 \\ \hline 7 \\ \hline \end{array}, \begin{array}{|c|c|c|c|c|c|} \hline 4 & 4 & 4 & 4 & 4 & 4 \\ \hline \end{array} \right).$$

One can check that  $I$  is a well-defined, weight-preserving, sign-reversing involution on  $X$ .

Let  $Z$  be the set of fixed points of  $I$ . We see from the description of  $I$  that  $Z$  consists of all triples  $(n - 1, T, \boxed{j})$  where  $j$  does not appear in  $T$ . All of these triples have sign  $(-1)^{n-1}$ . The proof will be complete if we can find a weight-preserving bijection  $g : Z \rightarrow Y$ . We define  $g$  by inserting a marked copy of  $j$  into its proper position in the increasing sequence  $T$ . The inverse map takes an increasing sequence of size  $n$  with one marked element and removes the marked element. For example,

$$g \left( 4, \begin{array}{|c|} \hline 2 \\ \hline 5 \\ \hline 7 \\ \hline 8 \\ \hline \end{array}, \begin{array}{|c|} \hline 3 \\ \hline \end{array} \right) = \begin{array}{|c|} \hline 2 \\ \hline 3^* \\ \hline 5 \\ \hline 7 \\ \hline 8 \\ \hline \end{array}.$$

□

## 10.20 Power-Sum Expansion of $h_n$ and $e_n$

We can use the recursions in 10.93 and 10.94 to compute expansions for  $h_n$  and  $e_n$  in terms of the power-sum symmetric polynomials  $p_\mu$ .

**10.95. Example.** We know that  $h_0 = 1$  and  $h_1 = p_1$ . Next, since  $h_0 p_2 + h_1 p_1 = 2h_2$ , we

find that  $h_2 = (p_{(2)} + p_{(1,1)})/2$ . For  $n = 3$ , we have

$$h_0p_3 + h_1p_2 + h_2p_1 = 3h_3,$$

so that

$$h_3 = \frac{1}{3} \left( p_3 + p_1p_2 + \left[ \frac{p_2 + p_1^2}{2} \right] p_1 \right) = (1/3)p_{(3)} + (1/2)p_{(2,1)} + (1/6)p_{(1,1,1)}.$$

For  $n = 4$ , we use the relation

$$h_0p_4 + h_1p_3 + h_2p_2 + h_3p_1 = 4h_4$$

to find, after some calculations,

$$h_4 = (1/4)p_{(4)} + (1/3)p_{(3,1)} + (1/8)p_{(2,2)} + (1/4)p_{(2,1,1)} + (1/24)p_{(1,1,1,1)}.$$

These formulas become nicer if we multiply through by  $n!$ . For instance,

$$\begin{aligned} 3!h_3 &= 2p_{(3)} + 3p_{(2,1)} + 1p_{(1,1,1)}; \\ 4!h_4 &= 6p_{(4)} + 8p_{(3,1)} + 3p_{(2,2)} + 6p_{(2,1,1)} + 1p_{(1,1,1,1)}. \end{aligned}$$

Similar formulas can be derived for  $n!e_n$ , but here some signs occur. For instance, calculations with (10.9) lead to the identities

$$\begin{aligned} 3!e_3 &= 2p_{(3)} - 3p_{(2,1)} + 1p_{(1,1,1)}; \\ 4!e_4 &= -6p_{(4)} + 8p_{(3,1)} + 3p_{(2,2)} - 6p_{(2,1,1)} + 1p_{(1,1,1,1)}. \end{aligned}$$

By comparing the coefficients in the power-sum expansion of  $4!h_4$  to the entries in Table 9.1, the reader may be led to conjecture the following result.

**10.96. Theorem: Power-Sum Expansion of  $h_n$ .** For all  $n, N \geq 1$ , the following identity is valid in  $\Lambda_N$ :

$$n!h_n = \sum_{\mu \in \text{Par}(n)} (n!/z_\mu) p_\mu. \quad (10.10)$$

*Proof.* Recall from 9.134 that  $n!/z_\mu$  is the number of permutations  $\sigma \in S_n$  with cycle type  $\mu$ . This suggests the following combinatorial interpretations for the two sides of (10.10). The left side counts all pairs  $(w, T)$ , where  $w = w_1w_2 \cdots w_n \in S_n$  is a permutation written in *one-line form* and  $T = (i_1 \leq i_2 \leq \cdots \leq i_n)$  is an element of  $\text{SSYT}_N((n))$ . Let  $X$  be the set of all such pairs, weighted by the content of  $T$ . For example, here is a typical element of  $X$  when  $n = 8$ , written as a two-rowed array:

$$z_0 = \left[ \begin{array}{cccccccc} w: & 4 & 2 & 5 & 8 & 3 & 7 & 1 & 6 \\ T: & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 \end{array} \right].$$

The right side of (10.10) counts all triples  $(\mu, \sigma, C)$ , where  $\mu \in \text{Par}(n)$ ,  $\sigma \in S_n$  is a permutation with cycle type  $\mu$ , and  $C : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, N\}$  is a *coloring* of the numbers  $1, \dots, n$  using  $N$  available colors such that all elements in the same cycle of  $\sigma$  are assigned the same color (cf. §9.19). Let the *weight* of such a triple be  $\prod_{k=1}^n x_{C(k)}$ , and let  $Y$  be the set of all such weighted triples. For example, a typical element of  $Y$  is the triple

$$y_0 = \left( (3, 2, 2, 1), (1, 6, 3)(2, 5)(7, 4)(8), \left( \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 2 & 3 & 3 & 2 & 3 & 3 & 3 \end{array} \right) \right).$$

To see why the factor  $p_\mu(x_1, \dots, x_N)$  arises, consider how we may choose the coloring function  $C$  once  $\mu$  and  $\sigma$  have been selected. Now, we know  $\sigma$  is a product of cycles of lengths  $\mu_1, \mu_2, \dots, \mu_l$ . Choose the common color of the elements in the first cycle in any of  $N$  ways. Since  $\mu_1$  elements all receive the same color, the generating function for this choice is  $x_1^{\mu_1} + x_2^{\mu_1} + \dots + x_N^{\mu_1} = p_{\mu_1}(x_1, \dots, x_N)$ . Next, choose the common color of the elements in the second cycle, which gives a factor of  $p_{\mu_2}$ , and so on. Multiplying the generating functions for these choices gives  $p_\mu(x_1, \dots, x_N)$ .

To complete the proof, we define weight-preserving maps  $f : Y \rightarrow X$  and  $g : X \rightarrow Y$  that are inverses of each other. To understand the definition of  $f$ , recall that a given  $\sigma \in S_n$  can be written in cycle notation in several different ways, since the cycles can be presented in any order, and elements within each cycle can be cyclically permuted. Given  $(\mu, \sigma, C)$ , we will specify one particular cycle notation for  $\sigma$  that depends on  $C$ , as follows. First, cycles colored with smaller colors are to be written before cycles colored with larger colors. Second, elements within each cycle are cyclically shifted so that the first element in each cycle is the smallest element appearing in that cycle. Third, if there are several cycles that have the same color, these cycles are ordered so that their minimal elements decrease from left to right. For example, starting with the object  $y_0$  above, we obtain the following cycle notation for  $\sigma$ :  $(2, 5)(8)(4, 7)(1, 6, 3)$ . Note that  $(2, 5)$  is written first because this cycle has color 2. The other cycles, which are all colored 3, are presented in the given order because  $8 > 4 > 1$ . Finally, to compute  $f((\mu, \sigma, C))$ , we *erase the parentheses* from the chosen cycle notation for  $\sigma$  and write the color  $C(i)$  directly beneath each  $i$  in the resulting word. For example,

$$f(y_0) = \begin{bmatrix} w : & 2 & 5 & 8 & 4 & 7 & 1 & 6 & 3 \\ T : & 2 & 2 & 3 & 3 & 3 & 3 & 3 & 3 \end{bmatrix}.$$

One may check that  $f$  is well-defined, maps into  $X$ , and preserves weights.

Now consider how to define the inverse map  $g : X \rightarrow Y$ . Given  $(w, T) \in X$  with  $w = w_1 \cdots w_n$  and  $T = i_1 \leq \dots \leq i_n$ , the coloring map  $C$  is defined by setting  $C(w_j) = i_j$  for all  $j$ . To recover  $\sigma$  from  $w$  and  $T$ , we need to add parentheses to  $w$  to recreate the cycle notation satisfying the rules above. For each color  $i$  in turn, look at the substring of  $w$  consisting of the symbols located above the  $i$ 's in  $T$ . Scan this substring from left to right, and begin a new cycle each time a number is encountered that is smaller than all of the preceding numbers in this substring. (The numbers that begin new cycles will be called *left-to-right minima relative to color  $i$* .) This procedure defines  $\sigma$ , and finally we set  $\mu = \text{type}(\sigma)$ . For example,

$$g(z_0) = \left( (2, 2, 1, 1, 1, 1), (4)(2, 5)(8)(3, 7)(1)(6), \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 1 & 2 & 1 & 1 & 3 & 2 & 2 \end{pmatrix} \right).$$

The reader should check that  $g(f(y_0)) = y_0$  and  $f(g(z_0)) = z_0$ . One can similarly verify that  $g \circ f = \text{id}_Y$  and  $f \circ g = \text{id}_X$ , so the proof is complete.  $\square$

Before considering the analogous theorem for  $e_n$ , we introduce the following notation.

**10.97. Definition: The Sign Factor  $\epsilon_\mu$ .** For every partition  $\mu \vdash n$ , let

$$\epsilon_\mu = (-1)^{n-\ell(\mu)} = \prod_{i=1}^{\ell(\mu)} (-1)^{\mu_i-1}.$$

We proved in 9.34 that  $\epsilon_\mu = \text{sgn}(\sigma)$  for all  $\sigma \in S_n$  such that  $\text{type}(\sigma) = \mu$ .

**10.98. Theorem: Power-Sum Expansion of  $e_n$ .** For all  $n, N \geq 1$ , the following identity is valid in  $\Lambda_N$ :

$$n!e_n = \sum_{\mu \in \text{Par}(n)} \epsilon_\mu (n!/z_\mu) p_\mu. \quad (10.11)$$

*Proof.* We use the notation  $X, Y, f, g, z_0$ , and  $y_0$  from the proof of 10.96. Recall that  $\sum_{\mu \vdash n} (n!/z_\mu) p_\mu$  is the generating function for the weighted set  $Y$ . To model the right side of (10.11), we need to get the sign factors  $\epsilon_\mu$  into this sum. We accomplish this by assigning *signs* to objects in  $Y$  as follows. Given  $(\mu, \sigma, C) \in Y$ , write  $\sigma$  in cycle notation as described previously. Attach a  $+$  to the first (minimal) element of each cycle, and attach a  $-$  to the remaining elements in each cycle. The overall sign of  $(\mu, \sigma, C)$  is the product of these signs, which is  $\prod_i (-1)^{\mu_i - 1} = \epsilon_\mu$ . For example, the object  $y_0$  considered previously is now written

$$y_0 = \left( (3, 2, 2, 1), (2^+, 5^-)(8^+)(4^+, 7^-)(1^+, 6^-, 3^-), \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 2 & 3 & 3 & 2 & 3 & 3 & 3 \end{pmatrix} \right);$$

the sign of this object is  $(-1)^4 = +1$ .

The next step is to transfer these signs to the objects in  $X$  using the weight-preserving bijection  $f : Y \rightarrow X$ . Given  $(w, T) \in X$ , find the left-to-right minima relative to each color  $i$  (as discussed in the definition of  $g$  in the proof of 10.96). Attach a  $+$  to these numbers and a  $-$  to all other numbers in  $w$ . For example,  $f(y_0)$  is now written

$$f(y_0) = \left[ \begin{array}{l} w : \quad 2^+ \quad 5^- \quad 8^+ \quad 4^+ \quad 7^- \quad 1^+ \quad 6^- \quad 3^- \\ T : \quad 2 \quad 2 \quad 3 \quad 3 \quad 3 \quad 3 \quad 3 \quad 3 \end{array} \right].$$

As another example,

$$z_0 = \left[ \begin{array}{l} w : \quad 4^+ \quad 2^+ \quad 5^- \quad 8^+ \quad 3^+ \quad 7^- \quad 1^+ \quad 6^+ \\ T : \quad 1 \quad 1 \quad 1 \quad 2 \quad 2 \quad 2 \quad 2 \quad 3 \end{array} \right].$$

The bijections  $f$  and  $g$  now preserve both signs and weights, by the way we defined signs of objects in  $X$ . It follows that  $\sum_{(w, T) \in X} \text{sgn}((w, T)) \text{wt}((w, T))$  is precisely the right side of (10.11).

Now we define a sign-reversing, weight-preserving involution  $I : X \rightarrow X$ . Fix  $(w, T) \in X$ . If all the entries of  $T$  are distinct, then  $(w, T)$  will be a fixed point of  $I$ . We observe at once that such a fixed point is necessarily positive, and the generating function for such objects is precisely  $n!e_n(x_1, \dots, x_N)$ . On the other hand, suppose some color  $i$  appears more than one time in  $T$ . Choose the smallest color  $i$  with this property, and let  $w_k, w_{k+1}$  be the first two symbols in the substring of  $w$  located above this color. Define  $I((w, T))$  by switching  $w_k$  and  $w_{k+1}$ ; one checks that this is a weight-preserving involution. Furthermore, one may verify that switching these two symbols will change the number of left-to-right minima (relative to color  $i$ ) by exactly 1. For example,

$$I(f(y_0)) = \left[ \begin{array}{l} w : \quad 5^+ \quad 2^+ \quad 8^+ \quad 4^+ \quad 7^- \quad 1^+ \quad 6^- \quad 3^- \\ T : \quad 2 \quad 2 \quad 3 \quad 3 \quad 3 \quad 3 \quad 3 \quad 3 \end{array} \right].$$

As another example,

$$I(z_0) = \left[ \begin{array}{l} w : \quad 2^+ \quad 4^- \quad 5^- \quad 8^+ \quad 3^+ \quad 7^- \quad 1^+ \quad 6^+ \\ T : \quad 1 \quad 1 \quad 1 \quad 2 \quad 2 \quad 2 \quad 2 \quad 3 \end{array} \right].$$

In general, note that  $w_k$  is always labeled by  $+$ ;  $w_{k+1}$  is labeled  $+$  iff  $w_k > w_{k+1}$ ; and the signs attached to numbers following  $w_{k+1}$  do not depend on the order of the two symbols  $w_k, w_{k+1}$ . We have now shown that  $I$  is sign-reversing, so the proof is complete.  $\square$

## 10.21 The Involution $\omega$

Recall from §10.16 that  $p_1, \dots, p_N \in \Lambda_N$  are algebraically independent over  $K$ , so that we can view  $\Lambda_N$  as the polynomial ring  $K[p_1, \dots, p_N]$ . By the universal mapping property



for polynomial rings (see 7.102), we can define a ring homomorphism with domain  $\Lambda_N$  by sending each  $p_i$  to an arbitrary element of a given ring containing  $K$ . We now apply this technique to define a certain homomorphism on  $\Lambda_N$ .

**10.99. Definition: The Map  $\omega$ .** Let  $\omega : \Lambda_N \rightarrow \Lambda_N$  be the unique  $K$ -linear ring homomorphism such that  $\omega(p_i) = (-1)^{i-1}p_i$  for  $1 \leq i \leq N$ .

**10.100. Theorem:  $\omega$  is an Involution.** We have  $\omega^2 = \text{id}_{\Lambda_N}$ ; in particular,  $\omega$  is an automorphism of  $\Lambda_N$ .

*Proof.* Observe that  $\omega^2(p_i) = \omega(\omega(p_i)) = \omega((-1)^{i-1}p_i) = (-1)^{i-1}(-1)^{i-1}p_i = p_i = \text{id}(p_i)$  for  $1 \leq i \leq N$ . Since  $\omega^2$  and  $\text{id}$  are ring homomorphisms on  $\Lambda_N$  that fix each  $c \in K$  and have the same effect on every  $p_i$ , the uniqueness part of the UMP for polynomial rings shows that  $\omega^2 = \text{id}$ . Since  $\omega$  has a two-sided inverse (namely, itself),  $\omega$  is a bijection.  $\square$

Let us investigate the effect of  $\omega$  on various bases of  $\Lambda_N$ .

**10.101. Theorem: Action of  $\omega$  on  $p$ 's,  $h$ 's, and  $e$ 's.** Suppose  $\nu \in \text{Par}_N(k)'$ , so  $\nu_i \leq N$  for all  $i$ . The following identities hold in  $\Lambda_N$ : (a)  $\omega(p_\nu) = \epsilon_\nu p_\nu$ ; (b)  $\omega(h_\nu) = e_\nu$ ; (c)  $\omega(e_\nu) = h_\nu$ .

*Proof.* (a) Since  $\omega$  is a ring homomorphism and  $\nu_i \leq N$  for all  $i$ ,

$$\omega(p_\nu) = \omega\left(\prod_{i=1}^{\ell(\nu)} p_{\nu_i}\right) = \prod_{i=1}^{\ell(\nu)} \omega(p_{\nu_i}) = \prod_{i=1}^{\ell(\nu)} (-1)^{\nu_i-1} p_{\nu_i} = \epsilon_\nu p_\nu.$$

(b) First, for any  $n \leq N$ , we have

$$\omega(h_n) = \omega\left(\sum_{\mu \in \text{Par}(n)} z_\mu^{-1} p_\mu\right) = \sum_{\mu \in \text{Par}(n)} z_\mu^{-1} \omega(p_\mu) = \sum_{\mu \in \text{Par}(n)} \epsilon_\mu z_\mu^{-1} p_\mu = e_n$$

by 10.96 and 10.98. Since  $\omega$  preserves multiplication,  $\omega(h_\nu) = e_\nu$  follows.

(c) Part (c) follows by applying  $\omega$  to both sides of (b), since  $\omega^2 = \text{id}$ .  $\square$

**10.102. Theorem: Action of  $\omega$  on  $s_\lambda$ .** If  $\lambda \in \text{Par}(n)$  and  $n \leq N$ , then  $\omega(s_\lambda) = s_{\lambda'}$  in  $\Lambda_N$ .

*Proof.* From 10.69, we know that for each  $\mu \vdash n$ ,

$$h_\mu(x_1, \dots, x_N) = \sum_{\lambda \in \text{Par}(n)} K_{\lambda, \mu} s_\lambda(x_1, \dots, x_N).$$

We can combine these equations into a single vector equation  $\mathbf{H} = \mathbf{K}^t \mathbf{S}$  where  $\mathbf{H} = (h_\mu : \mu \in \text{Par}(n))$  and  $\mathbf{S} = (s_\lambda : \lambda \in \text{Par}(n))$ . Since  $\mathbf{K}^t$  (the transpose of the Kostka matrix) is unitriangular and hence invertible,  $\mathbf{S} = (\mathbf{K}^t)^{-1} \mathbf{H}$  is the *unique* vector  $\mathbf{v}$  satisfying  $\mathbf{H} = \mathbf{K}^t \mathbf{v}$ .

From 10.75, we know that for each  $\mu \vdash n$ ,

$$e_\mu(x_1, \dots, x_N) = \sum_{\lambda \in \text{Par}(n)} K_{\lambda, \mu} s_{\lambda'}(x_1, \dots, x_N).$$

Applying the linear map  $\omega$  to these equations produces the equations

$$h_\mu = \sum_{\lambda \in \text{Par}(n)} K_{\lambda, \mu} \omega(s_{\lambda'}).$$

But this says that the vector  $\mathbf{v} = (\omega(s_{\lambda'}) : \lambda \in \text{Par}(n))$  satisfies  $\mathbf{H} = \mathbf{K}^t \mathbf{v}$ . By the uniqueness property mentioned above,  $\mathbf{v} = \mathbf{S}$ . So, for all  $\lambda \in \text{Par}(n)$ ,  $s_\lambda = \omega(s_{\lambda'})$ . Replacing  $\lambda$  by  $\lambda'$  (or applying  $\omega$  to both sides) gives the result.  $\square$

What happens if we apply  $\omega$  to the monomial basis of  $\Lambda_N$ ? Since  $\omega$  is a  $K$ -linear bijection, we get another basis of  $\Lambda_N$  that is different from those discussed so far. This basis is hard to describe directly, so it is given the following name.

**10.103. Definition: Forgotten Basis for  $\Lambda_N$ .** For each  $\lambda \in \text{Par}_N$ , define the *forgotten symmetric polynomial*  $\text{fgt}_\lambda = \omega(m_\lambda)$ . The set  $\{\text{fgt}_\lambda : \lambda \in \text{Par}_N(k)\}$  is a basis of  $\Lambda_N^k$ .

## 10.22 Permutations and Tableaux

Iteration of the tableau insertion algorithm (§10.9) leads to some remarkable bijections that map permutations, words, and matrices to certain pairs of tableaux. These bijections were studied by Robinson, Schensted, and Knuth, and are therefore called *RSK correspondences*. We begin in this section by showing how permutations can be encoded using pairs of standard tableaux of the same shape.

**10.104. Theorem: RSK Correspondence for Permutations.** There is a bijection  $\text{RSK} : S_n \rightarrow \bigcup_{\lambda \in \text{Par}(n)} \text{SYT}(\lambda) \times \text{SYT}(\lambda)$ . Given  $\text{RSK}(w) = (P(w), Q(w))$ , we call  $P(w)$  the *insertion tableau for  $w$*  and  $Q(w)$  the *recording tableau for  $w$* .

*Proof.* Let  $w \in S_n$  have one-line form  $w = w_1 w_2 \cdots w_n$ . We construct a sequence of tableaux  $P_0, P_1, \dots, P_n = P(w)$  and a sequence of tableaux  $Q_0, Q_1, \dots, Q_n = Q(w)$  as follows. Initially, let  $P_0$  and  $Q_0$  be empty tableaux of shape  $(0)$ . Suppose  $1 \leq i \leq n$  and  $P_{i-1}, Q_{i-1}$  have already been constructed. Define  $P_i = P_{i-1} \leftarrow w_i$  (the tableau obtained by insertion of  $w_i$  into  $P_{i-1}$ ). Let  $(a, b)$  be the new cell in  $P_i$  created by this insertion. Define  $Q_i$  to be the tableau obtained from  $Q_{i-1}$  by placing the value  $i$  in the new cell  $(a, b)$ . Informally, we build  $P(w)$  by inserting  $w_1, \dots, w_n$  (in this order) into an initially empty tableau. We build  $Q(w)$  by placing the numbers  $1, 2, \dots, n$  (in this order) in the new boxes created by each insertion. By construction,  $Q(w)$  has the same shape as  $P(w)$ . Furthermore, since the new box at each stage is a corner box, one sees that  $Q(w)$  will be a standard tableau. Finally, set  $\text{RSK}(w) = (P(w), Q(w))$ .

To see that RSK is a bijection, we present an algorithm for computing the inverse map. Let  $(P, Q)$  be any pair of standard tableaux of the same shape  $\lambda \in \text{Par}(n)$ . The idea is to recover the one-line form  $w_1 \cdots w_n$  in reverse by uninserting entries from  $P$ , using the entries in  $Q$  to decide which box to remove at each stage (cf. §10.10). To begin, note that  $n$  occurs in some corner box  $(a, b)$  of  $Q$  (since  $Q$  is standard). Apply reverse insertion to  $P$  starting at  $(a, b)$  to obtain the unique tableau  $P_{n-1}$  and value  $w_n$  such that  $P_{n-1} \leftarrow w_n$  is  $P$  with new box  $(a, b)$  (see 10.60). Let  $Q_{n-1}$  be the tableau obtained by erasing  $n$  from  $Q$ . Continue similarly: having computed  $P_i$  and  $Q_i$  such that  $Q_i$  is a standard tableau with  $i$  cells, let  $(a, b)$  be the corner box of  $Q_i$  containing  $i$ . Apply reverse insertion to  $P_i$  starting at  $(a, b)$  to obtain  $P_{i-1}$  and  $w_i$ . Then delete  $i$  from  $Q_i$  to obtain a standard tableau  $Q_{i-1}$  with  $i-1$  cells. The resulting word  $w = w_1 w_2 \cdots w_n$  is a permutation of  $\{1, 2, \dots, n\}$  (since  $P$  contains each of these values exactly once), and our argument has shown that  $w$  is the *unique* object satisfying  $\text{RSK}(w) = (P, Q)$ . So RSK is a bijection.  $\square$

**10.105. Example.** Let  $w = 35164872 \in S_8$ . Figure 10.1 illustrates the computation of  $\text{RSK}(w) = (P(w), Q(w))$ . As an example of the inverse computation, let us determine the permutation  $v = \text{RSK}^{-1}(Q(w), P(w))$  (note that we have switched the order of the insertion and recording tableaux). Figure 10.2 displays the reverse insertions used to find  $v_n, v_{n-1}, \dots, v_1$ . We see that  $v = 38152476$ .

	Insertion Tableau	Recording Tableau																								
insert 3:	<table><tr><td>3</td></tr></table>	3	<table><tr><td>1</td></tr></table>	1																						
3																										
1																										
insert 5:	<table><tr><td>3</td><td>5</td></tr></table>	3	5	<table><tr><td>1</td><td>2</td></tr></table>	1	2																				
3	5																									
1	2																									
insert 1:	<table><tr><td>1</td><td>5</td></tr><tr><td>3</td><td></td></tr></table>	1	5	3		<table><tr><td>1</td><td>2</td></tr><tr><td>3</td><td></td></tr></table>	1	2	3																	
1	5																									
3																										
1	2																									
3																										
insert 6:	<table><tr><td>1</td><td>5</td><td>6</td></tr><tr><td>3</td><td></td><td></td></tr></table>	1	5	6	3			<table><tr><td>1</td><td>2</td><td>4</td></tr><tr><td>3</td><td></td><td></td></tr></table>	1	2	4	3														
1	5	6																								
3																										
1	2	4																								
3																										
insert 4:	<table><tr><td>1</td><td>4</td><td>6</td></tr><tr><td>3</td><td>5</td><td></td></tr></table>	1	4	6	3	5		<table><tr><td>1</td><td>2</td><td>4</td></tr><tr><td>3</td><td>5</td><td></td></tr></table>	1	2	4	3	5													
1	4	6																								
3	5																									
1	2	4																								
3	5																									
insert 8:	<table><tr><td>1</td><td>4</td><td>6</td><td>8</td></tr><tr><td>3</td><td>5</td><td></td><td></td></tr></table>	1	4	6	8	3	5			<table><tr><td>1</td><td>2</td><td>4</td><td>6</td></tr><tr><td>3</td><td>5</td><td></td><td></td></tr></table>	1	2	4	6	3	5										
1	4	6	8																							
3	5																									
1	2	4	6																							
3	5																									
insert 7:	<table><tr><td>1</td><td>4</td><td>6</td><td>7</td></tr><tr><td>3</td><td>5</td><td>8</td><td></td></tr></table>	1	4	6	7	3	5	8		<table><tr><td>1</td><td>2</td><td>4</td><td>6</td></tr><tr><td>3</td><td>5</td><td>7</td><td></td></tr></table>	1	2	4	6	3	5	7									
1	4	6	7																							
3	5	8																								
1	2	4	6																							
3	5	7																								
insert 2:	<table><tr><td>1</td><td>2</td><td>6</td><td>7</td></tr><tr><td>3</td><td>4</td><td>8</td><td></td></tr><tr><td>5</td><td></td><td></td><td></td></tr></table>	1	2	6	7	3	4	8		5				<table><tr><td>1</td><td>2</td><td>4</td><td>6</td></tr><tr><td>3</td><td>5</td><td>7</td><td></td></tr><tr><td>8</td><td></td><td></td><td></td></tr></table>	1	2	4	6	3	5	7		8			
1	2	6	7																							
3	4	8																								
5																										
1	2	4	6																							
3	5	7																								
8																										

**FIGURE 10.1**  
Computation of  $\text{RSK}(35164872)$ .

Let us compare the two-line forms of  $w$  and  $v$ :

$$w = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 5 & 1 & 6 & 4 & 8 & 7 & 2 \end{pmatrix}; \quad v = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 8 & 1 & 5 & 2 & 4 & 7 & 6 \end{pmatrix}.$$

We see that  $v$  and  $w$  are inverse permutations!

The phenomenon observed in the last example holds in general: if  $w \mapsto (P, Q)$  under the RSK correspondence, then  $w^{-1} \mapsto (Q, P)$ . To prove this fact, we must introduce a new way of visualizing the construction of the insertion and recording tableaux for  $w$ .

**10.106. Definition: Cartesian Graph of a Permutation.** Given a permutation  $w = w_1w_2 \cdots w_n \in S_n$ , the graph of  $w$  (in the  $xy$ -plane) is the set  $G(w) = \{(i, w_i) : 1 \leq i \leq n\}$ .

For example, the graph of  $w = 35164872$  is drawn in Figure 10.3.

To analyze the creation of the insertion and recording tableaux for  $\text{RSK}(w)$ , we will annotate the graph of  $w$  by drawing lines as described in the following definitions.

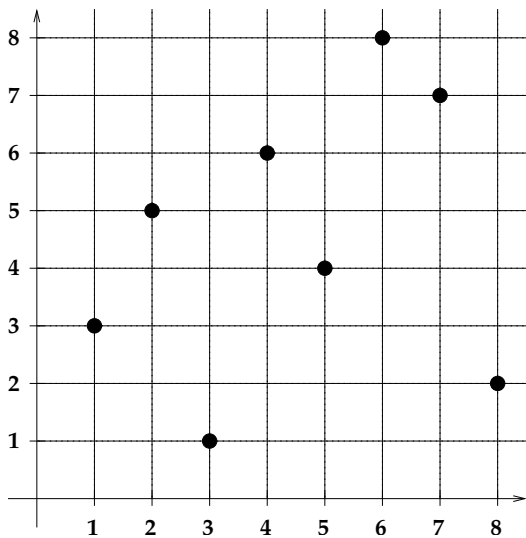
**10.107. Definition: Shadow Lines.** Let  $S = \{(x_1, y_1), \dots, (x_k, y_k)\}$  be a finite set of points in the first quadrant. The *shadow* of  $S$  is

$$\text{Shd}(S) = \{(u, v) \in \mathbb{R}^2 : \text{for some } i, u \geq x_i \text{ and } v \geq y_i\}.$$

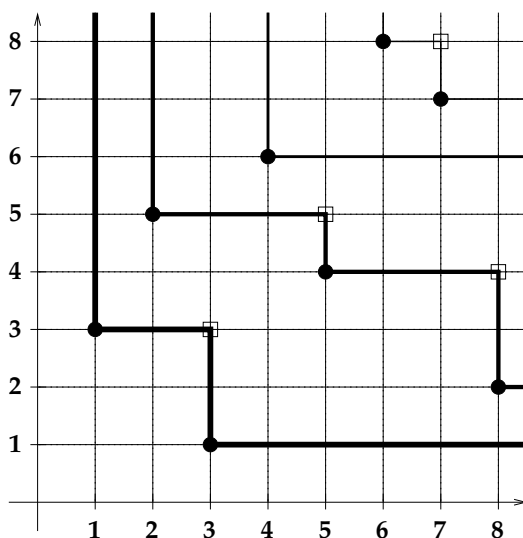
Informally, the shadow consists of all points northeast of some point in  $S$ . The *first shadow line*  $L_1(S)$  is the boundary of  $\text{Shd}(S)$ . This boundary consists of an infinite vertical ray (part of the line  $x = a_1$ , say), followed by zero or more alternating horizontal and vertical line segments, followed by an infinite horizontal ray (part of the line  $y = b_1$ , say). Call  $a_1$  and  $b_1$  the *x-coordinate* and *y-coordinate* associated to this shadow line. Next, let  $S_1$  be the set of points in  $S$  that lie on the first shadow line of  $S$ . The *second shadow line*  $L_2(S)$  is

	Insertion Tableau	Recording Tableau	Output Value																								
initial tableau:	<table><tr><td>1</td><td>2</td><td>4</td><td>6</td></tr><tr><td>3</td><td>5</td><td><u>7</u></td><td></td></tr><tr><td>8</td><td></td><td></td><td></td></tr></table>	1	2	4	6	3	5	<u>7</u>		8				<table><tr><td>1</td><td>2</td><td>6</td><td>7</td></tr><tr><td>3</td><td>4</td><td><u>8</u></td><td></td></tr><tr><td>5</td><td></td><td></td><td></td></tr></table>	1	2	6	7	3	4	<u>8</u>		5				
1	2	4	6																								
3	5	<u>7</u>																									
8																											
1	2	6	7																								
3	4	<u>8</u>																									
5																											
uninsert 7:	<table><tr><td>1</td><td>2</td><td>4</td><td><u>7</u></td></tr><tr><td>3</td><td>5</td><td></td><td></td></tr><tr><td>8</td><td></td><td></td><td></td></tr></table>	1	2	4	<u>7</u>	3	5			8				<table><tr><td>1</td><td>2</td><td>6</td><td><u>7</u></td></tr><tr><td>3</td><td>4</td><td></td><td></td></tr><tr><td>5</td><td></td><td></td><td></td></tr></table>	1	2	6	<u>7</u>	3	4			5				6
1	2	4	<u>7</u>																								
3	5																										
8																											
1	2	6	<u>7</u>																								
3	4																										
5																											
uninsert 7:	<table><tr><td>1</td><td>2</td><td><u>4</u></td><td></td></tr><tr><td>3</td><td>5</td><td></td><td></td></tr><tr><td>8</td><td></td><td></td><td></td></tr></table>	1	2	<u>4</u>		3	5			8				<table><tr><td>1</td><td>2</td><td><u>6</u></td><td></td></tr><tr><td>3</td><td>4</td><td></td><td></td></tr><tr><td>5</td><td></td><td></td><td></td></tr></table>	1	2	<u>6</u>		3	4			5				7
1	2	<u>4</u>																									
3	5																										
8																											
1	2	<u>6</u>																									
3	4																										
5																											
uninsert 4:	<table><tr><td>1</td><td>2</td><td></td><td></td></tr><tr><td>3</td><td>5</td><td></td><td></td></tr><tr><td><u>8</u></td><td></td><td></td><td></td></tr></table>	1	2			3	5			<u>8</u>				<table><tr><td>1</td><td>2</td><td></td><td></td></tr><tr><td>3</td><td>4</td><td></td><td></td></tr><tr><td><u>5</u></td><td></td><td></td><td></td></tr></table>	1	2			3	4			<u>5</u>				4
1	2																										
3	5																										
<u>8</u>																											
1	2																										
3	4																										
<u>5</u>																											
uninsert 8:	<table><tr><td>1</td><td>5</td><td></td><td></td></tr><tr><td>3</td><td><u>8</u></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	1	5			3	<u>8</u>							<table><tr><td>1</td><td>2</td><td></td><td></td></tr><tr><td>3</td><td><u>4</u></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	1	2			3	<u>4</u>							2
1	5																										
3	<u>8</u>																										
1	2																										
3	<u>4</u>																										
uninsert 8:	<table><tr><td>1</td><td>8</td><td></td><td></td></tr><tr><td><u>3</u></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	1	8			<u>3</u>								<table><tr><td>1</td><td>2</td><td></td><td></td></tr><tr><td><u>3</u></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	1	2			<u>3</u>								5
1	8																										
<u>3</u>																											
1	2																										
<u>3</u>																											
uninsert 3:	<table><tr><td>3</td><td><u>8</u></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	3	<u>8</u>											<table><tr><td>1</td><td><u>2</u></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	1	<u>2</u>											1
3	<u>8</u>																										
1	<u>2</u>																										
uninsert 8:	<table><tr><td><u>3</u></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	<u>3</u>												<table><tr><td><u>1</u></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>	<u>1</u>												8
<u>3</u>																											
<u>1</u>																											
uninsert 3:	empty	empty	3																								

**FIGURE 10.2**  
Mapping pairs of standard tableaux to permutations.



**FIGURE 10.3**  
Cartesian graph of a permutation.

**FIGURE 10.4**

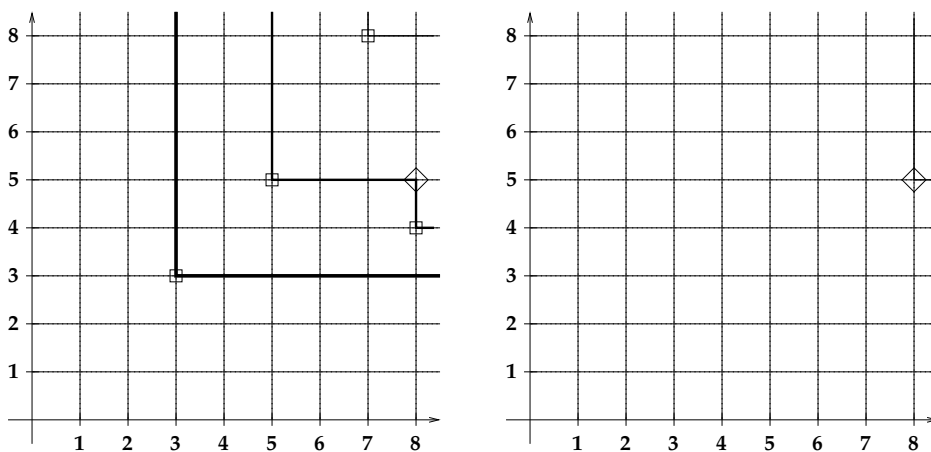
Shadow lines for a permutation graph.

the boundary of  $\text{Shd}(S \sim S_1)$ , which has associated coordinates  $(a_2, b_2)$ . Letting  $S_2$  be the points in  $S$  that lie on the second shadow line, the *third shadow line*  $L_3(S)$  is the boundary of  $\text{Shd}(S \sim (S_1 \cup S_2))$ . We continue to generate shadow lines in this way until all points of  $S$  lie on some shadow line. Finally, the *first-order shadow diagram* of  $w \in S_n$  consists of all shadow lines associated to the graph  $G(w)$ .

**10.108. Example.** The first-order shadow diagram of  $w = 35164872$  is drawn in Figure 10.4. The  $x$ -coordinates associated to the shadow lines of  $w$  are 1, 2, 4, 6. These  $x$ -coordinates agree with the entries in the first row of the recording tableau  $Q(w)$ , which we computed in 10.105. Similarly, the  $y$ -coordinates of the shadow lines are 1, 2, 6, 7, which are precisely the entries in the first row of the insertion tableau  $P(w)$ . The next result explains why this happens, and shows that the shadow diagram contains complete information about the evolution of the first rows of  $P(w)$  and  $Q(w)$ .

**10.109. Theorem: Shadow Lines and RSK.** Let  $w \in S_n$  have first-order shadow lines  $L_1, \dots, L_k$  with associated coordinates  $(x_1, y_1), \dots, (x_k, y_k)$ . Let  $P_0, P_1, \dots, P_n = P(w)$  and  $Q_0, Q_1, \dots, Q_n = Q(w)$  be the sequences of tableaux generated in the computation of  $\text{RSK}(w)$ . For  $0 \leq i \leq n$ , the  $y$ -coordinates of the intersections of the shadow lines with the line  $x = i + (1/2)$  are the entries in the first row of  $P_i$ , whereas the entries in the first row of  $Q_i$  consist of all  $x_j \leq i$ . Whenever some shadow line  $L_r$  has a vertical segment from  $(i, a)$  down to  $(i, b)$ , then  $b = w_i$  and the insertion  $P_i = P_{i-1} \leftarrow w_i$  bumps the value  $a$  out of the  $r$ th cell in the first row of  $P_{i-1}$ .

*Proof.* We proceed by induction on  $i \geq 0$ . The theorem holds when  $i = 0$ , since  $P_0$  and  $Q_0$  are empty, and no shadow lines intersect the line  $x = 1/2$ . Assume the result holds for  $i - 1 < n$ . Then the first row of  $P_{i-1}$  is  $a_1 < a_2 < \dots < a_j$ , and the  $a$ 's are the  $y$ -coordinates where the shadow lines hit the line  $x = i - 1/2$ . Consider the point  $(i, w_i)$ , which is the unique point in  $G(w)$  on the line  $x = i$ . First consider the case  $w_i > a_j$ . In this case, the first  $j$  shadow lines all pass underneath  $(i, w_i)$ . It follows that  $(i, w_i)$  is the first point of

**FIGURE 10.5**

Higher-order shadow diagrams.

$G(w)$  on shadow line  $L_{j+1}(G(w))$ , so  $x_{j+1} = i$ . When we insert  $w_i$  into  $P_{i-1}$ ,  $w_i$  goes at the end of the first row of  $P_{i-1}$  (since it exceeds the last entry  $a_j$ ), and we place  $i$  in the corresponding cell in the first row of  $Q_i$ . The statements in the theorem regarding  $P_i$  and  $Q_i$  are true in this case. Now consider the case  $w_i < a_j$ . Suppose  $a_r$  is the smallest value in the first row of  $P_{i-1}$  exceeding  $w_i$ . Then insertion of  $w_i$  into  $P_{i-1}$  bumps  $a_r$  out of the first row. On the other hand, the point  $(i, w_i)$  lies between the points  $(i, a_{r-1})$  and  $(i, a_r)$  in the shadow diagram (taking  $a_0 = 0$ ). It follows from the way the shadow lines are drawn that shadow line  $L_r$  must drop from  $(i, a_r)$  to  $(i, w_i)$  when it reaches the line  $x = i$ . The statements of the theorem therefore hold for  $i$  in this case as well.  $\square$

To analyze the rows of  $P(w)$  and  $Q(w)$  below the first row, we iterate the shadow diagram construction as follows.

**10.110. Definition: Iterated Shadow Diagrams.** Let  $L_1, \dots, L_k$  be the shadow lines associated to a given subset  $S$  of  $\mathbb{R}^2$ . An *inner corner* is a point  $(a, b)$  at the top of one of the vertical segments of some shadow line. Let  $S'$  be the set of inner corners associated to  $S$ . The *second-order shadow diagram* of  $S$  is the shadow diagram associated to  $S'$ . We iterate this process to define all higher-order shadow diagrams of  $S$ .

For example, taking  $w = 35164872$ , Figure 10.5 displays the second-order and third-order shadow diagrams for  $G(w)$ .

**10.111. Theorem: Higher-Order Shadows and RSK.** For  $w \in S_n$ , let  $L_1, \dots, L_k$  be the shadow lines in the  $r$ th-order shadow diagram for  $w \in S_n$ , with associated coordinates  $(x_1, y_1), \dots, (x_k, y_k)$ . Let  $P_0, P_1, \dots, P_n = P(w)$  and  $Q_0, Q_1, \dots, Q_n = Q(w)$  be the sequences of tableaux generated in the computation of  $\text{RSK}(w)$ . For  $0 \leq i \leq n$ , the  $y$ -coordinates of the intersections of the shadow lines with the line  $x = i + (1/2)$  are the entries in the  $r$ th row of  $P_i$ , whereas the entries in the  $r$ th row of  $Q_i$  consist of all  $x_j \leq i$ . Whenever some shadow line  $L_c$  has a vertical segment from  $(i, a)$  down to  $(i, b)$ , then  $b$  is the value bumped out of row  $r - 1$  by the insertion  $P_i = P_{i-1} \leftarrow w_i$ , and  $b$  bumps the value  $a$  out of the  $c$ th cell in row  $r$  of  $P_{i-1}$ . (Take  $b = w_i$  when  $r = 1$ .)

*Proof.* We use induction on  $r \geq 1$ . The base case  $r = 1$  was proved in 10.109. Consider  $r = 2$

next. The proof of 10.109 shows that the inner corners of the first-order shadow diagram of  $w$  are precisely those points  $(i, b)$  such that  $b$  is bumped out of the first row of  $P_{i-1}$  and inserted into the second row of  $P_{i-1}$  when forming  $P_i$ . The argument used to prove 10.109 can now be applied to this set of points. Whenever a point  $(i, b)$  lies above all second-order shadow lines approaching the line  $x = i$  from the left,  $b$  gets inserted in a new cell at the end of the second row of  $P_i$  and the corresponding cell in  $Q_i$  receives the label  $i$ . Otherwise, if  $(i, b)$  lies between shadow lines  $L_{c-1}$  and  $L_c$  in the second-order diagram, then  $b$  bumps the value in the  $c$ th cell of the second row of  $P_{i-1}$  into the third row, and shadow line  $L_c$  moves down to level  $b$  when it reaches  $x = i$ . The statements in the theorem (for  $r = 2$ ) follow exactly as before by induction on  $i$ . Iterating this argument establishes the analogous results for each  $r > 2$ .  $\square$

**10.112. Theorem: RSK and Inversion.** For all  $w \in S_n$ , if  $\text{RSK}(w) = (P, Q)$ , then  $\text{RSK}(w^{-1}) = (Q, P)$ .

*Proof.* Consider the picture consisting of  $G(w)$  and the first-order shadow diagram. Suppose the shadow lines have associated  $x$ -coordinates  $(a_1, \dots, a_k)$  and  $y$ -coordinates  $(b_1, \dots, b_k)$ . Let us reflect the picture through the line  $y = x$  (which interchanges  $x$ -coordinates and  $y$ -coordinates). This reflection changes  $G(w)$  into  $G(w^{-1})$ , since  $(x, y) \in G(w)$  iff  $y = w(x)$  iff  $x = w^{-1}(y)$  iff  $(y, x) \in G(w^{-1})$ . We see from the geometric definition that the shadow lines for  $w$  get reflected into the shadow lines for  $w^{-1}$ . It follows from 10.109 that the first row of both  $Q(w)$  and  $P(w^{-1})$  is  $a_1, \dots, a_k$ , whereas the first row of both  $P(w)$  and  $Q(w^{-1})$  is  $b_1, \dots, b_k$ . The inner corners for  $w^{-1}$  are the reflections of the inner corners for  $w$ . So, we can apply the same argument to the higher-order shadow diagrams of  $w$  and  $w^{-1}$ . It follows that all rows of  $P(w^{-1})$  (resp.  $Q(w^{-1})$ ) agree with the corresponding rows of  $Q(w)$  (resp.  $P(w)$ ).  $\square$

## 10.23 Words and Tableaux

We now generalize the RSK algorithm to operate on arbitrary words, not just permutations.

**10.113. Theorem: RSK Correspondence for Words.** Let  $W = X^n$  be the set of  $n$ -letter words over an ordered alphabet  $X$ . There is a bijection

$$\text{RSK} : W \rightarrow \bigcup_{\lambda \in \text{Par}(n)} \text{SSYT}_X(\lambda) \times \text{SYT}(\lambda).$$

We write  $\text{RSK}(w) = (P(w), Q(w))$ ;  $P(w)$  is the *insertion tableau* for  $w$  and  $Q(w)$  is the *recording tableau* for  $w$ . For all  $x \in X$ , the number of  $x$ 's in  $w$  is the same as the number of  $x$ 's in  $P(w)$ .

*Proof.* Given  $w = w_1 w_2 \cdots w_n \in W$ , we define sequences of tableaux  $P_0, P_1, \dots, P_n$  and  $Q_0, Q_1, \dots, Q_n$  as follows.  $P_0$  and  $Q_0$  are the empty tableau. If  $P_{i-1}$  and  $Q_{i-1}$  have been computed for some  $i \leq n$ , let  $P_i = P_{i-1} \leftarrow w_i$ . Suppose this insertion creates a new box  $(c, d)$ ; then we form  $Q_i$  from  $Q_{i-1}$  by placing the value  $i$  in the box  $(c, d)$ . By induction on  $i$ , we see that every  $P_i$  is semistandard with values in  $X$ , every  $Q_i$  is standard, and  $P_i$  and  $Q_i$  have the same shape. We set  $\text{RSK}(w) = (P_n, Q_n)$ . The letters in  $P_n$  (counting repetitions) are exactly the letters in  $w$ , so the last statement of the theorem holds.

Next we describe the inverse algorithm. Given  $(P, Q)$  with  $P$  semistandard and  $Q$  standard of the same shape, we construct semistandard tableaux  $P_n, P_{n-1}, \dots, P_0$ , standard

	Insertion Tableau	Recording Tableau																
insert 2:	<table><tr><td>2</td></tr></table>	2	<table><tr><td>1</td></tr></table>	1														
2																		
1																		
insert 1:	<table><tr><td>1</td></tr><tr><td>2</td></tr></table>	1	2	<table><tr><td>1</td></tr><tr><td>2</td></tr></table>	1	2												
1																		
2																		
1																		
2																		
insert 1:	<table><tr><td>1</td><td>1</td></tr><tr><td>2</td></tr></table>	1	1	2	<table><tr><td>1</td><td>3</td></tr><tr><td>2</td></tr></table>	1	3	2										
1	1																	
2																		
1	3																	
2																		
insert 3:	<table><tr><td>1</td><td>1</td><td>3</td></tr><tr><td>2</td></tr></table>	1	1	3	2	<table><tr><td>1</td><td>3</td><td>4</td></tr><tr><td>2</td></tr></table>	1	3	4	2								
1	1	3																
2																		
1	3	4																
2																		
insert 2:	<table><tr><td>1</td><td>1</td><td>2</td></tr><tr><td>2</td><td>3</td></tr></table>	1	1	2	2	3	<table><tr><td>1</td><td>3</td><td>4</td></tr><tr><td>2</td><td>5</td></tr></table>	1	3	4	2	5						
1	1	2																
2	3																	
1	3	4																
2	5																	
insert 1:	<table><tr><td>1</td><td>1</td><td>1</td></tr><tr><td>2</td><td>2</td></tr><tr><td>3</td></tr></table>	1	1	1	2	2	3	<table><tr><td>1</td><td>3</td><td>4</td></tr><tr><td>2</td><td>5</td></tr><tr><td>6</td></tr></table>	1	3	4	2	5	6				
1	1	1																
2	2																	
3																		
1	3	4																
2	5																	
6																		
insert 3:	<table><tr><td>1</td><td>1</td><td>1</td><td>3</td></tr><tr><td>2</td><td>2</td></tr><tr><td>3</td></tr></table>	1	1	1	3	2	2	3	<table><tr><td>1</td><td>3</td><td>4</td><td>7</td></tr><tr><td>2</td><td>5</td></tr><tr><td>6</td></tr></table>	1	3	4	7	2	5	6		
1	1	1	3															
2	2																	
3																		
1	3	4	7															
2	5																	
6																		
insert 1:	<table><tr><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>2</td><td>2</td><td>3</td></tr><tr><td>3</td></tr></table>	1	1	1	1	2	2	3	3	<table><tr><td>1</td><td>3</td><td>4</td><td>7</td></tr><tr><td>2</td><td>5</td><td>8</td></tr><tr><td>6</td></tr></table>	1	3	4	7	2	5	8	6
1	1	1	1															
2	2	3																
3																		
1	3	4	7															
2	5	8																
6																		

**FIGURE 10.6**  
Computation of  $\text{RSK}(21132131)$ .

tableaux  $Q_n, Q_{n-1}, \dots, Q_0$ , and letters  $w_n, w_{n-1}, \dots, w_1$  as follows. Initially,  $P_n = P$  and  $Q_n = Q$ . Suppose, for some  $i \leq n$ , that we have already constructed tableaux  $P_i$  and  $Q_i$  such that these tableaux have the same shape and consist of  $i$  boxes,  $P_i$  is semistandard, and  $Q_i$  is standard. The value  $i$  lies in a corner cell of  $Q_i$ ; perform uninsertion starting from the same cell in  $P_i$  to get a smaller semistandard tableau  $P_{i-1}$  and a letter  $w_i$ . Let  $Q_{i-1}$  be  $Q_i$  with the  $i$  erased. At the end, output the word  $w_1 w_2 \cdots w_n$ . Using 10.60 and induction, one checks that  $w = w_1 \cdots w_n$  is the unique word  $w$  with  $\text{RSK}(w) = (P, Q)$ . So the RSK algorithm is a bijection.  $\square$

**10.114. Example.** Let  $w = 21132131$ . We compute  $\text{RSK}(w)$  in Figure 10.6.

Next we investigate how the RSK algorithm is related to certain statistics on words and tableaux.

**10.115. Definition: Descents and Major Index for Standard Tableaux.** Let  $Q$  be a standard tableau with  $n$  cells. The *descent set of  $Q$* , denoted  $\text{Des}(Q)$ , is the set of all  $k < n$  such that  $k+1$  appears in a lower row of  $Q$  than  $k$ . The *descent count of  $Q$* , denoted  $\text{des}(Q)$ , is  $|\text{Des}(Q)|$ . The *major index of  $Q$* , denoted  $\text{maj}(Q)$ , is  $\sum_{k \in \text{Des}(Q)} k$ . (Compare to 6.27, which gives the analogous definitions for words.)

**10.116. Example.** For the standard tableau  $Q = Q(w)$  shown at the bottom of Figure 10.6, we have  $\text{Des}(Q) = \{1, 4, 5, 7\}$ ,  $\text{des}(Q) = 4$ , and  $\text{maj}(Q) = 17$ . Here,  $w = 21132131$ . Note that  $\text{Des}(w) = \{1, 4, 5, 7\}$ ,  $\text{des}(w) = 4$ , and  $\text{maj}(w) = 17$ . This is not a coincidence.

**10.117. Theorem: RSK Preserves Descents and Major Index.** Let  $w \in X^n$  be a



word with recording tableau  $Q = Q(w)$ . Then  $\text{Des}(w) = \text{Des}(Q)$ ,  $\text{des}(w) = \text{des}(Q)$ , and  $\text{maj}(w) = \text{maj}(Q)$ .

*Proof.* It suffices to prove  $\text{Des}(w) = \text{Des}(Q)$ . Let  $P_0, P_1, \dots, P_n$  and  $Q_0, Q_1, \dots, Q_n = Q$  be the sequences of tableaux computed when we apply the RSK algorithm to  $w$ . For each  $k < n$ , note that  $k \in \text{Des}(w)$  iff  $w_k > w_{k+1}$ , whereas  $k \in \text{Des}(Q)$  iff  $k+1$  appears in a row below  $k$ 's row in  $Q$ . So, for each  $k < n$ , we must prove  $w_k > w_{k+1}$  iff  $k+1$  appears in a row below  $k$ 's row in  $Q$ . For this, we use the bumping comparison theorem 10.62. Consider the double insertion  $(P_{k-1} \leftarrow w_k) \leftarrow w_{k+1}$ . Let the new box in  $P_{k-1} \leftarrow w_k$  be  $(i, j)$ , and let the new box in  $(P_{k-1} \leftarrow w_k) \leftarrow w_{k+1}$  be  $(r, s)$ . By definition of the recording tableau,  $Q(i, j) = k$  and  $Q(r, s) = k+1$ . Now, if  $w_k > w_{k+1}$ , part 2 of 10.62 says that  $i < r$  (and  $j \geq s$ ). So  $k+1$  appears in a lower row than  $k$  in  $Q$ . If instead  $w_k \leq w_{k+1}$ , part 1 of 10.62 says that  $i \geq r$  (and  $j < s$ ). So  $k+1$  does not appear in a lower row than  $k$  in  $Q$ .  $\square$

Now suppose the letters  $x_1, \dots, x_N$  in  $X$  are variables in some polynomial ring. Then we can view a word  $w = w_1 \dots w_n \in X^n$  as a *monomial* in the  $x_j$ 's by forming the product of all the letters appearing in  $w$  (counting repetitions). Using 2.9, we can write

$$\sum_{w \in X^n} w = (x_1 + \dots + x_N)^n = p_{(1^n)}(x_1, \dots, x_N). \quad (10.12)$$

We can use the RSK algorithm and 10.117 to obtain a related identity involving Schur polynomials.

**10.118. Theorem: Schur Expansion for Words weighted by Major Index.** Let  $X = \{x_1, \dots, x_N\}$  where the  $x_i$ 's are commuting indeterminates ordered by  $x_1 < x_2 < \dots < x_N$ . For every  $n \geq 1$ ,

$$\sum_{w \in X^n} t^{\text{maj}(w)} w = \sum_{\lambda \in \text{Par}(n)} \left( \sum_{Q \in \text{SYT}(\lambda)} t^{\text{maj}(Q)} \right) s_\lambda(x_1, \dots, x_N). \quad (10.13)$$

*Proof.* The left side of (10.13) is the generating function for the weighted set  $X^n$ , where the weight of a word  $w$  is  $t^{\text{maj}(w)}$ . On the other hand, let us weight each set  $\text{SSYT}_X(\lambda)$  by taking  $\text{wt}(P)$  to be the product of the entries in  $P$ . Comparing to the definition of Schur polynomials, we see that the generating function for this weighted set is  $s_\lambda(x_1, \dots, x_N)$ . Next, weight each set  $\text{SYT}(\lambda)$  by taking  $\text{wt}(Q) = t^{\text{maj}(Q)}$  for  $Q \in \text{SYT}(\lambda)$ . Finally, consider the set  $\bigcup_{\lambda \in \text{Par}(n)} \text{SSYT}_X(\lambda) \times \text{SYT}(\lambda)$  weighted by setting  $\text{wt}(P, Q) = \text{wt}(P) \text{wt}(Q)$ . By the sum and product rules for weighted sets, the generating function for this last weighted set is precisely the right side of (10.13). To complete the proof, note that the RSK map is a weight-preserving bijection between  $X^n$  and  $\bigcup_{\lambda \in \text{Par}(n)} \text{SSYT}_X(\lambda) \times \text{SYT}(\lambda)$ , because of 10.113 and 10.117.  $\square$

Setting  $t = 1$  in the previous result and using (10.12) gives the following formula for  $p_{(1^n)}$  in terms of Schur polynomials.

**10.119. Theorem: Schur Expansion of  $p_{(1^n)}$ .** For all  $n, N \geq 1$ ,

$$p_{(1^n)}(x_1, \dots, x_N) = \sum_{\lambda \in \text{Par}(n)} |\text{SYT}(\lambda)| s_\lambda(x_1, \dots, x_N).$$

**10.120. Remark.** The RSK correspondence can also be used to find the length of the longest weakly increasing or strictly decreasing subsequence of a given word. For details, see §12.11.

## 10.24 Matrices and Tableaux

Performing the RSK map on a word produces a pair consisting of one *semistandard* tableau and one *standard* tableau. We now define an RSK operation on matrices that will map each matrix to a pair of semistandard tableaux of the same shape. The first step is to encode the matrix as a certain biword.

**10.121. Definition: Biword of a Matrix.** Let  $A = (a_{ij})$  be an  $M \times N$  matrix with entries in  $\mathbb{N}$ . The *biword* of  $A$  is a two-row array

$$\text{bw}(A) = \begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}$$

constructed as follows. Start with the empty biword, and scan the rows of  $A$  from top to bottom, reading each row from left to right. Whenever a nonzero integer  $a_{ij}$  is encountered in the scan, write down  $a_{ij}$  copies of  $\begin{pmatrix} i \\ j \end{pmatrix}$  at the end of the current biword. The top row of  $\text{bw}(A)$  is called the *row word* of  $A$  and denoted  $r(A)$ . The bottom row of  $\text{bw}(A)$  is called the *column word* of  $A$  and denoted  $c(A)$ .

**10.122. Example.** Suppose  $A$  is the matrix

$$A = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The biword of  $A$  is

$$\text{bw}(A) = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 2 & 3 \\ 1 & 1 & 3 & 2 & 4 & 4 & 4 & 3 \end{pmatrix}.$$

**10.123. Theorem: Matrices vs. Biwords.** Let  $X$  be the set of all  $M \times N$  matrices with entries in  $\mathbb{N}$ . Let  $Y$  be the set of all biwords  $w = \begin{pmatrix} i_1 & i_2 & \cdots & i_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}$  satisfying the following conditions: (a)  $i_1 \leq i_2 \leq \cdots \leq i_k$ ; (b) if  $i_s = i_{s+1}$ , then  $j_s \leq j_{s+1}$ ; (c)  $i_s \leq M$  for all  $s$ ; (d)  $j_s \leq N$  for all  $s$ . The map  $\text{bw} : X \rightarrow Y$  is a bijection. Suppose  $A = (a_{ij})$  has biword  $\text{bw}(A)$  as above. Then  $i$  appears  $\sum_j a_{ij}$  times in  $r(A)$ ,  $j$  appears  $\sum_i a_{ij}$  times in  $c(A)$ , and  $k = \sum_{i,j} a_{ij}$ .

*Proof.* To show that  $\text{bw}$  maps  $X$  into  $Y$ , we must show that  $\text{bw}(A)$  satisfies conditions (a) through (d). Condition (a) holds since we scan the rows of  $A$  from top to bottom. Condition (b) holds since each row is scanned from left to right. Condition (c) holds since  $A$  has  $M$  rows. Condition (d) holds since  $A$  has  $N$  columns. We can invert  $\text{bw}$  as follows. Given a biword  $w \in Y$ , let  $A$  be the  $M \times N$  matrix such that, for all  $i \leq M$  and all  $j \leq N$ ,  $a_{ij}$  is the number of indices  $s$  with  $i_s = i$  and  $j_s = j$ . The last statements in the theorem follow from the way we constructed  $r(A)$  and  $c(A)$ .  $\square$

**10.124. Theorem: RSK Correspondence for Biwords.** Let  $Y$  be the set of biwords defined in 10.123. Let  $Z$  be the set of pairs  $(P, Q)$  of semistandard tableaux of the same shape such that  $P$  has entries in  $\{1, 2, \dots, N\}$  and  $Q$  has entries in  $\{1, 2, \dots, M\}$ . There is a bijection  $\text{RSK} : Y \rightarrow Z$  that “preserves content.” This means that if  $\begin{pmatrix} v \\ w \end{pmatrix} \in Y$  maps to  $(P, Q) \in Z$ , then for all  $i \leq M$ ,  $v$  and  $Q$  contain the same number of  $i$ ’s, and for all  $j \leq N$ ,  $w$  and  $P$  contain the same number of  $j$ ’s.

*Proof.* Write  $v = i_1 \leq i_2 \leq \cdots \leq i_k$  and  $w = j_1, j_2, \dots, j_k$ , where  $i_s = i_{s+1}$  implies  $j_s \leq j_{s+1}$ . As in the previous RSK maps, we build sequences of insertion tableaux  $P_0, P_1, \dots, P_k$  and recording tableaux  $Q_0, Q_1, \dots, Q_k$ . Initially,  $P_0$  and  $Q_0$  are empty. Having constructed  $P_s$  and  $Q_s$ , let  $P_{s+1} = P_s \leftarrow j_{s+1}$ . If the new box created by this insertion is  $(a, b)$ , obtain  $Q_{s+1}$  from  $Q_s$  by setting  $Q_{s+1}(a, b) = i_{s+1}$ . The final output is  $\text{RSK}(\binom{v}{w}) = (P_k, Q_k)$ .

By construction,  $P_k$  is semistandard with entries consisting of the letters in  $w$ , and the entries of  $Q_k$  are the letters in  $v$ . But, is  $Q = Q_k$  a semistandard tableau? To see that it is, note that we obtain  $Q$  by successively placing a weakly increasing sequence of numbers  $i_1 \leq i_2 \leq \cdots \leq i_k$  into new corner boxes of an initially empty tableau. It follows that the rows and columns of  $Q$  weakly increase. To see that columns of  $Q$  *strictly* increase, consider what happens during the placement of a run of equal numbers into  $Q$ , say  $i = i_s = i_{s+1} = \cdots = i_t$ . By definition of  $Y$ , we have  $j_s \leq j_{s+1} \leq \cdots \leq j_t$ . When we insert this weakly increasing sequence into the  $P$ -tableau, the resulting sequence of new boxes forms a horizontal strip by 10.63. So, the corresponding boxes in  $Q$  (which consist of all the boxes labeled  $i$  in  $Q$ ) also form a horizontal strip. This means that there are never two equal numbers in a given column of  $Q$ .

The inverse algorithm reconstructs the words  $v$  and  $w$  in reverse, starting with  $i_k$  and  $j_k$ . Given  $(P, Q)$ , look for the rightmost occurrence of the largest letter in  $Q$ , which must reside in a corner box. Let  $i_k$  be this letter. Erase this cell from  $Q$ , and perform reverse insertion on  $P$  starting at the same cell to recover  $j_k$ . Iterate this process on the resulting smaller tableaux. We have  $i_k \geq \cdots \geq i_1$  since we remove the largest letter in  $Q$  at each stage. When we remove a string of equal letters from  $Q$ , say  $i = i_t = i_{t-1} = \cdots = i_s$ , the associated letters removed from  $P$  must satisfy  $j_t \geq j_{t-1} \geq \cdots \geq j_s$ . This follows, as above, from the bumping comparison theorem 10.62. For instance, if  $j_{t-1} > j_t$ , then the new box created at stage  $t$  would be weakly left of the new box created at stage  $t-1$ , which contradicts the requirement of choosing the *rightmost*  $i$  in  $Q$  when recovering  $i_t$  and  $j_t$ . It follows that the inverse algorithm does produce a biword in  $Y$ , as required.  $\square$

Composing the two preceding bijections gives the following result.

**10.125. Theorem: RSK Correspondence for Matrices.** For every  $M, N \geq 1$ , there is a bijection between the set of  $M \times N$  matrices with entries in  $\mathbb{N}$  and the set

$$\bigcup_{\lambda \in \text{Par}} \text{SSYT}_N(\lambda) \times \text{SSYT}_M(\lambda)$$

given by  $A \mapsto \text{RSK}(\text{bw}(A))$ . If  $(a_{ij}) \mapsto (P, Q)$  under this bijection, then the number of  $j$ 's in  $P$  is  $\sum_i a_{ij}$ , and the number of  $i$ 's in  $Q$  is  $\sum_j a_{ij}$ .

**10.126. Example.** Let us compute the pair of tableaux associated to the matrix  $A$  from 10.122. Looking at the biword of  $A$ , we must insert the sequence  $c(A) = (1, 1, 3, 2, 4, 4, 4, 3)$  into the  $P$ -tableau, recording the entries in  $r(A) = (1, 1, 1, 2, 2, 2, 2, 3)$  in the  $Q$ -tableau. This computation appears in Figure 10.7.

**10.127. Theorem: Cauchy Identity for Schur Polynomials.** For all  $M, N \geq 1$ , we have the formal power series identity in  $\mathbb{Q}[[x_1, \dots, x_M, y_1, \dots, y_N]]$ :

$$\prod_{i=1}^M \prod_{j=1}^N \frac{1}{1 - x_i y_j} = \sum_{\lambda \in \text{Par}} s_{\lambda}(y_1, \dots, y_N) s_{\lambda}(x_1, \dots, x_M). \quad (10.14)$$

*Proof.* We interpret each side as the generating function for a suitable set of weighted

	Insertion Tableau	Recording Tableau																								
insert 1, record 1:	<table><tr><td>1</td></tr></table>	1	<table><tr><td>1</td></tr></table>	1																						
1																										
1																										
insert 1, record 1:	<table><tr><td>1</td><td>1</td></tr></table>	1	1	<table><tr><td>1</td><td>1</td></tr></table>	1	1																				
1	1																									
1	1																									
insert 3, record 1:	<table><tr><td>1</td><td>1</td><td>3</td></tr></table>	1	1	3	<table><tr><td>1</td><td>1</td><td>1</td></tr></table>	1	1	1																		
1	1	3																								
1	1	1																								
insert 2, record 2:	<table><tr><td>1</td><td>1</td><td>2</td></tr><tr><td>3</td><td></td><td></td></tr></table>	1	1	2	3			<table><tr><td>1</td><td>1</td><td>1</td></tr><tr><td>2</td><td></td><td></td></tr></table>	1	1	1	2														
1	1	2																								
3																										
1	1	1																								
2																										
insert 4, record 2:	<table><tr><td>1</td><td>1</td><td>2</td><td>4</td></tr><tr><td>3</td><td></td><td></td><td></td></tr></table>	1	1	2	4	3				<table><tr><td>1</td><td>1</td><td>1</td><td>2</td></tr><tr><td>2</td><td></td><td></td><td></td></tr></table>	1	1	1	2	2											
1	1	2	4																							
3																										
1	1	1	2																							
2																										
insert 4, record 2:	<table><tr><td>1</td><td>1</td><td>2</td><td>4</td><td>4</td></tr><tr><td>3</td><td></td><td></td><td></td><td></td></tr></table>	1	1	2	4	4	3					<table><tr><td>1</td><td>1</td><td>1</td><td>2</td><td>2</td></tr><tr><td>2</td><td></td><td></td><td></td><td></td></tr></table>	1	1	1	2	2	2								
1	1	2	4	4																						
3																										
1	1	1	2	2																						
2																										
insert 4, record 2:	<table><tr><td>1</td><td>1</td><td>2</td><td>4</td><td>4</td><td>4</td></tr><tr><td>3</td><td></td><td></td><td></td><td></td><td></td></tr></table>	1	1	2	4	4	4	3						<table><tr><td>1</td><td>1</td><td>1</td><td>2</td><td>2</td><td>2</td></tr><tr><td>2</td><td></td><td></td><td></td><td></td><td></td></tr></table>	1	1	1	2	2	2	2					
1	1	2	4	4	4																					
3																										
1	1	1	2	2	2																					
2																										
insert 3, record 3:	<table><tr><td>1</td><td>1</td><td>2</td><td>3</td><td>4</td><td>4</td></tr><tr><td>3</td><td>3</td><td></td><td></td><td></td><td></td></tr></table>	1	1	2	3	4	4	3	3					<table><tr><td>1</td><td>1</td><td>1</td><td>2</td><td>2</td><td>2</td></tr><tr><td>2</td><td>3</td><td></td><td></td><td></td><td></td></tr></table>	1	1	1	2	2	2	2	3				
1	1	2	3	4	4																					
3	3																									
1	1	1	2	2	2																					
2	3																									

**FIGURE 10.7**  
Applying the RSK map to a biword.

objects. For the left side, consider  $M \times N$  matrices with entries in  $\mathbb{N}$ . Let the weight of a matrix  $A = (a_{ij})$  be

$$\text{wt}(A) = \prod_{i=1}^M \prod_{j=1}^N (x_i y_j)^{a_{ij}}.$$

We can build such a matrix by choosing the entries  $a_{ij} \in \mathbb{N}$  one at a time. For fixed  $i$  and  $j$ , the generating function for the choice of  $a_{ij}$  is

$$1 + x_i y_j + (x_i y_j)^2 + \cdots + (x_i y_j)^k + \cdots = \frac{1}{1 - x_i y_j}.$$

By the product rule for weighted sets, we see that the left side of (10.14) is the generating function for this set of matrices. On the other hand, the RSK bijection converts each matrix in this set to a pair of semistandard tableaux of the same shape. This bijection  $A \mapsto (P, Q)$  will be weight-preserving provided that we weight each occurrence of  $j$  in  $P$  by  $y_j$  and each occurrence of  $i$  in  $Q$  by  $x_i$ . With these weights, the generating function for  $\text{SSYT}_N(\lambda)$  is  $s_\lambda(y_1, \dots, y_N)$ , and the generating function for  $\text{SSYT}_M(\lambda)$  is  $s_\lambda(x_1, \dots, x_M)$ . It now follows from the sum and product rules for weighted sets that the right side of (10.14) is the generating function for the weighted set  $\bigcup_{\lambda \in \text{Par}} \text{SSYT}_N(\lambda) \times \text{SSYT}_M(\lambda)$ . Since RSK is a weight-preserving bijection, the proof is complete.  $\square$

**10.25 Cauchy Identities**

In the last section, we found a formula expressing the product  $\prod_{i,j}(1 - x_i y_j)^{-1}$  as a sum of products of Schur polynomials. Next, we derive other formulas for this product that involve other kinds of symmetric polynomials.

**10.128. Theorem: Cauchy Identities.** For all  $M, N \geq 1$ , we have the formal power series identities:

$$\begin{aligned} \prod_{i=1}^M \prod_{j=1}^N \frac{1}{1 - x_i y_j} &= \sum_{\lambda \in \text{Par}_N} h_{\lambda}(x_1, \dots, x_M) m_{\lambda}(y_1, \dots, y_N) \\ &= \sum_{\lambda \in \text{Par}_M} m_{\lambda}(x_1, \dots, x_M) h_{\lambda}(y_1, \dots, y_N) \\ &= \sum_{\lambda \in \text{Par}} \frac{p_{\lambda}(x_1, \dots, x_M) p_{\lambda}(y_1, \dots, y_N)}{z_{\lambda}}. \end{aligned}$$

*Proof.* Recall from 10.92 the product expansion

$$\prod_{i=1}^M \frac{1}{1 - x_i t} = \sum_{k=0}^{\infty} h_k(x_1, \dots, x_M) t^k.$$

Replacing  $t$  by  $y_j$ , where  $j$  is a fixed index, we obtain

$$\prod_{i=1}^M \frac{1}{1 - x_i y_j} = \sum_{k=0}^{\infty} h_k(x_1, \dots, x_M) y_j^k.$$

Taking the product over  $j$  gives

$$\prod_{i=1}^M \prod_{j=1}^N \frac{1}{1 - x_i y_j} = \prod_{j=1}^N \sum_{k_j=0}^{\infty} h_{k_j}(x_1, \dots, x_M) y_j^{k_j}.$$

Let us expand the product on the right side using the generalized distributive law (§2.1), suitably extended to handle infinite series within the product. We obtain

$$\prod_{i=1}^M \prod_{j=1}^N \frac{1}{1 - x_i y_j} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \prod_{j=1}^N h_{k_j}(x_1, \dots, x_M) y_j^{k_j}.$$

Let us reorganize the sum on the right side by grouping together summands indexed by sequences  $(k_1, \dots, k_N)$  that can be sorted to give the same partition  $\lambda$ . Since  $h_{k_1} h_{k_2} \cdots h_{k_N} = h_{\lambda}$  for all such sequences, the right side becomes

$$\sum_{\lambda \in \text{Par}_N} h_{\lambda}(x_1, \dots, x_M) \sum_{\substack{(k_1, \dots, k_N) \in \mathbb{N}^N: \\ \text{sort}(k_1, \dots, k_N) = \lambda}} y_1^{k_1} y_2^{k_2} \cdots y_N^{k_N}.$$

The inner sum is precisely the definition of  $m_{\lambda}(y_1, \dots, y_N)$ . So the first formula of the theorem is proved. The second formula follows from the same argument, interchanging the roles of the  $x$ 's and  $y$ 's.

To obtain the formula involving power sums, we again start with 10.92, which can be written

$$\prod_{k=1}^{MN} \frac{1}{1 - z_k t} = \sum_{n=0}^{\infty} h_n(z_1, \dots, z_{MN}) t^n.$$

Replace the  $MN$  variables  $z_k$  by the  $MN$  quantities  $x_i y_j$  (with  $1 \leq i \leq M$  and  $1 \leq j \leq N$ ). We obtain

$$\prod_{i=1}^M \prod_{j=1}^N \frac{1}{1 - x_i y_j t} = \sum_{n=0}^{\infty} h_n(x_1 y_1, x_1 y_2, \dots, x_M y_N) t^n.$$

Now use 10.96 to rewrite the right side in terms of power sums:

$$\prod_{i=1}^M \prod_{j=1}^N \frac{1}{1 - x_i y_j t} = \sum_{n=0}^{\infty} t^n \sum_{\lambda \in \text{Par}(n)} p_{\lambda}(x_1 y_1, x_1 y_2, \dots, x_M y_N) / z_{\lambda}.$$

Observe next that, for all  $k \geq 1$ ,

$$\begin{aligned} p_k(x_1 y_1, \dots, x_M y_N) &= \sum_{i=1}^M \sum_{j=1}^N (x_i y_j)^k = \sum_{i=1}^M \sum_{j=1}^N x_i^k y_j^k \\ &= \left( \sum_{i=1}^M x_i^k \right) \cdot \left( \sum_{j=1}^N y_j^k \right) = p_k(x_1, x_2, \dots, x_M) p_k(y_1, y_2, \dots, y_N). \end{aligned}$$

It follows from this that, for any partition  $\lambda$ ,

$$p_{\lambda}(x_1 y_1, \dots, x_M y_N) = p_{\lambda}(x_1, x_2, \dots, x_M) p_{\lambda}(y_1, y_2, \dots, y_N).$$

We therefore find that

$$\prod_{i=1}^M \prod_{j=1}^N \frac{1}{1 - x_i y_j t} = \sum_{n=0}^{\infty} t^n \sum_{\lambda \in \text{Par}(n)} \frac{p_{\lambda}(x_1, x_2, \dots, x_M) p_{\lambda}(y_1, y_2, \dots, y_N)}{z_{\lambda}}. \quad (10.15)$$

Setting  $t = 1$  gives the final formula of the theorem.  $\square$

## 10.26 Dual Bases

Now we introduce a scalar product on the vector spaces  $\Lambda_N^k$ . For convenience, we assume that  $N$  (the number of variables) is not less than  $k$  (the degree of the polynomials in the space), so that the various bases of  $\Lambda_N^k$  are indexed by *all* the integer partitions of  $k$ .

**10.129. Definition: Hall Scalar Product on  $\Lambda_N^k$ .** For  $N \geq k$ , define the *Hall scalar product* on the vector space  $\Lambda_N^k$  by setting (for all  $\mu, \nu \in \text{Par}(k)$ )

$$\langle p_{\mu}, p_{\nu} \rangle = 0 \text{ if } \mu \neq \nu, \quad \langle p_{\mu}, p_{\mu} \rangle = z_{\mu},$$

and extending by bilinearity. In more detail, given  $f, g \in \Lambda_N^k$ , choose scalars  $a_{\mu}, b_{\mu} \in K$  such that  $f = \sum_{\mu} a_{\mu} p_{\mu}$  and  $g = \sum_{\nu} b_{\nu} p_{\nu}$ . Then  $\langle f, g \rangle = \sum_{\mu} a_{\mu} b_{\mu} z_{\mu} \in K$ .

**10.130. Definition: Orthonormal Bases and Dual Bases.** Suppose  $N \geq k$  and  $B_1 = \{f_{\mu} : \mu \in \text{Par}(k)\}$  and  $B_2 = \{g_{\mu} : \mu \in \text{Par}(k)\}$  are two bases of  $\Lambda_N^k$ .  $B_1$  is called an *orthonormal basis* iff  $\langle f_{\mu}, f_{\nu} \rangle = \chi(\mu = \nu)$  for all  $\mu, \nu \in \text{Par}(k)$ .  $B_1$  and  $B_2$  are called *dual bases* iff  $\langle f_{\mu}, g_{\nu} \rangle = \chi(\mu = \nu)$  for all  $\mu, \nu \in \text{Par}(k)$ .

For example, taking  $F = \mathbb{C}$ ,  $\{p_{\mu} / \sqrt{z_{\mu}} : \mu \in \text{Par}(k)\}$  is an orthonormal basis of  $\Lambda_N^k$ . The next theorem allows us to detect dual bases by looking at expansions of the product  $\prod_{i,j} (1 - x_i y_j)^{-1}$ .

**10.131. Theorem: Characterization of Dual Bases.** Suppose  $N \geq k$  and  $B_1 = \{f_{\mu} : \mu \in \text{Par}(k)\}$  and  $B_2 = \{g_{\mu} : \mu \in \text{Par}(k)\}$  are two bases of  $\Lambda_N^k$ .  $B_1$  and  $B_2$  are dual bases iff

$$\left( \prod_{i=1}^N \prod_{j=1}^N \frac{1}{1 - x_i y_j t} \right)_k = \sum_{\mu \in \text{Par}(k)} f_{\mu}(x_1, \dots, x_N) g_{\mu}(y_1, \dots, y_N),$$

where the left side is the coefficient of  $t^k$  in the indicated product.

*Proof.* Comparing the displayed equation to (10.15), we must prove that  $B_1$  and  $B_2$  are dual bases iff

$$\sum_{\mu \in \text{Par}(k)} p_\mu(x_1, \dots, x_N) p_\mu(y_1, \dots, y_N) / z_\mu = \sum_{\mu \in \text{Par}(k)} f_\mu(x_1, \dots, x_N) g_\mu(y_1, \dots, y_N).$$

The idea of the proof is to convert each condition into a statement about matrices. Since  $\{p_\mu\}$  and  $\{p_\mu/z_\mu\}$  are bases of  $\Lambda_N^k$ , there exist scalars  $a_{\mu,\nu}, b_{\mu,\nu} \in K$  satisfying

$$f_\nu = \sum_{\mu} a_{\mu,\nu} p_\mu, \quad g_\nu = \sum_{\mu} b_{\mu,\nu} (p_\mu / z_\mu).$$

Define matrices  $\mathbf{A} = (a_{\mu,\nu})$  and  $\mathbf{B} = (b_{\mu,\nu})$ . By bilinearity, we compute (for all  $\lambda, \nu \in \text{Par}(k)$ ):

$$\begin{aligned} \langle f_\lambda, g_\nu \rangle &= \left\langle \sum_{\mu} a_{\mu,\lambda} p_\mu, \sum_{\rho} b_{\rho,\nu} p_\rho / z_\rho \right\rangle \\ &= \sum_{\mu, \rho} a_{\mu,\lambda} b_{\rho,\nu} \langle p_\mu, p_\rho / z_\rho \rangle \\ &= \sum_{\mu} a_{\mu,\lambda} b_{\mu,\nu} = (\mathbf{A}^t \mathbf{B})_{\lambda,\nu}. \end{aligned}$$

It follows that  $\{f_\lambda\}$  and  $\{g_\nu\}$  are dual bases iff  $\mathbf{A}^t \mathbf{B} = \mathbf{I}$  (the identity matrix of size  $|\text{Par}(k)|$ ).

On the other hand, writing  $\vec{x} = (x_1, \dots, x_N)$  and  $\vec{y} = (y_1, \dots, y_N)$ , we have

$$\sum_{\mu \in \text{Par}(k)} f_\mu(\vec{x}) g_\mu(\vec{y}) = \sum_{\mu, \alpha, \beta} a_{\alpha,\mu} b_{\beta,\mu} p_\alpha(\vec{x}) p_\beta(\vec{y}) / z_\beta.$$

Now, one may check that the polynomials

$$\{p_\alpha(\vec{x}) p_\beta(\vec{y}) / z_\beta : (\alpha, \beta) \in \text{Par}(k) \times \text{Par}(k)\}$$

are linearly independent, using the fact that the power-sum polynomials in one set of variables are linearly independent. It follows that the expression given above for  $\sum_{\mu \in \text{Par}(k)} f_\mu(\vec{x}) g_\mu(\vec{y})$  will be equal to  $\sum_{\alpha \in \text{Par}(k)} p_\alpha(\vec{x}) p_\alpha(\vec{y}) / z_\alpha$  iff  $\sum_{\mu} a_{\alpha,\mu} b_{\beta,\mu} = 0$  for all  $\alpha \neq \beta$  and  $\sum_{\mu} a_{\alpha,\mu} b_{\alpha,\mu} = 1$  for all  $\alpha$ . In matrix form, these equations say that  $\mathbf{A} \mathbf{B}^t = \mathbf{I}$ . This matrix equation is equivalent to  $\mathbf{B}^t \mathbf{A} = \mathbf{I}$  (since all the matrices are square), which is equivalent in turn to  $\mathbf{A}^t \mathbf{B} = \mathbf{I}$ . We saw above that this last condition holds iff  $B_1$  and  $B_2$  are dual bases, so the proof is complete.  $\square$

**10.132. Theorem: Dual Bases of  $\Lambda_N^k$ .** For  $N \geq k$ ,  $\{s_\mu(x_1, \dots, x_N) : \mu \in \text{Par}(k)\}$  is an orthonormal basis of  $\Lambda_N^k$ . Also  $\{m_\mu(x_1, \dots, x_N) : \mu \in \text{Par}(k)\}$  and  $\{h_\mu(x_1, \dots, x_N) : \mu \in \text{Par}(k)\}$  are dual bases of  $\Lambda_N^k$ .

*Proof.* In 10.127, replace every  $x_i$  by  $tx_i$ . Since  $s_\lambda$  is homogeneous of degree  $|\lambda|$ , we obtain

$$\prod_{i=1}^N \prod_{j=1}^N \frac{1}{1 - tx_i y_j} = \sum_{\lambda \in \text{Par}} s_\lambda(y_1, \dots, y_N) s_\lambda(x_1, \dots, x_N) t^{|\lambda|}.$$

Extracting the coefficient of  $t^k$  gives

$$\left( \prod_{i=1}^N \prod_{j=1}^N \frac{1}{1 - x_i y_j t} \right)_k = \sum_{\lambda \in \text{Par}(k)} s_\lambda(x_1, \dots, x_N) s_\lambda(y_1, \dots, y_N).$$

So 10.131 applies to show that  $\{s_\lambda : \lambda \in \text{Par}(k)\}$  is an orthonormal basis. We proceed similarly to see that the  $m$ 's and  $h$ 's are dual, starting with 10.128.  $\square$

**10.133. Theorem:  $\omega$  is an Isometry.** For  $N \geq k$ , the map  $\omega : \Lambda_N^k \rightarrow \Lambda_N^k$  is an isometry relative to the Hall scalar product. In other words, for all  $f, g \in \Lambda_N^k$ ,  $\langle \omega(f), \omega(g) \rangle = \langle f, g \rangle$ . Therefore,  $\omega$  sends an orthonormal basis (resp. dual bases) of  $\Lambda_N^k$  to an orthonormal basis (resp. dual bases) of  $\Lambda_N^k$ .

*Proof.* Write  $f = \sum_\mu a_\mu p_\mu$  and  $g = \sum_\nu b_\nu p_\nu$  for suitable scalars  $a_\mu, b_\nu \in K$ . By linearity of  $\omega$  and bilinearity of the Hall scalar product, we compute

$$\begin{aligned} \langle \omega(f), \omega(g) \rangle &= \left\langle \omega \left( \sum_\mu a_\mu p_\mu \right), \omega \left( \sum_\nu b_\nu p_\nu \right) \right\rangle \\ &= \sum_\mu \sum_\nu a_\mu b_\nu \langle \omega(p_\mu), \omega(p_\nu) \rangle \\ &= \sum_\mu \sum_\nu a_\mu b_\nu \epsilon_\mu \epsilon_\nu \langle p_\mu, p_\nu \rangle \\ &= \sum_\mu a_\mu b_\mu \epsilon_\mu^2 z_\mu. \end{aligned}$$

The last step follows since we only get a nonzero scalar product when  $\nu = \mu$ . Now, the last expression is

$$\sum_\mu a_\mu b_\mu z_\mu = \langle f, g \rangle. \quad \square$$

**10.134. Theorem: Duality of  $e_\mu$ 's and  $\text{fgt}_\nu$ 's.** For  $N \geq k$ , the bases  $\{e_\mu : \mu \in \text{Par}(k)\}$  and  $\{\text{fgt}_\mu : \mu \in \text{Par}(k)\}$  (forgotten basis) are dual. Moreover,

$$\left( \prod_{i=1}^N \prod_{j=1}^N \frac{1}{1 - x_i y_j t} \right)_k = \sum_{\lambda \in \text{Par}(k)} e_\lambda(x_1, \dots, x_N) \text{fgt}_\lambda(y_1, \dots, y_N).$$

*Proof.* We know that  $\{m_\mu\}$  and  $\{h_\mu\}$  are dual bases. Since  $\text{fgt}_\mu = \omega(m_\mu)$  and  $e_\mu = \omega(h_\mu)$ ,  $\{\text{fgt}_\mu\}$  and  $\{e_\mu\}$  are dual bases. The indicated product formula now follows from 10.131.  $\square$

## Summary

Table 10.2 summarizes information about five bases for the vector space  $\Lambda_N^k$  of symmetric polynomials in  $N$  variables that are homogeneous of degree  $k$ . The statements about dual bases assume  $N \geq k$ . Recall that  $\text{Par}_N(k)$  is the set of integer partitions of  $k$  into at most  $N$  parts, while  $\text{Par}_N(k)'$  is the set of partitions of  $k$  where every part is at most  $N$ . Table 10.3 gives formulas and recursions for expressing certain symmetric polynomials as linear combinations of other symmetric polynomials. Further identities of a similar kind can be found in the summary of Chapter 11.

- *Skew Shapes and Skew Schur Polynomials.* A skew shape  $\mu/\nu$  is obtained by removing the diagram of  $\nu$  from the diagram of  $\mu$ . A semistandard tableau of this shape is a filling



**TABLE 10.2**

Bases for symmetric polynomials.

Basis of $\Lambda_N^k$	Definition	Dual Basis	Action of $\omega$
Monomial $\{m_\mu : \mu \in \text{Par}_N(k)\}$	$m_\mu = \sum_{\substack{\alpha \in \mathbb{N}^N: \\ \text{sort}(\alpha) = \mu}} x_1^{\alpha_1} \cdots x_N^{\alpha_N}$	$\{h_\mu\}$	$\omega(m_\mu) = \text{fgt}_\mu$
Elementary $\{e_\mu : \mu \in \text{Par}_N(k)'\}$	$e_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq N} x_{i_1} x_{i_2} \cdots x_{i_k},$ $e_\mu = e_{\mu_1} e_{\mu_2} \cdots e_{\mu_s}$	$\{\text{fgt}_\mu\}$	$\omega(e_\mu) = h_\mu$
Complete $\{h_\mu : \mu \in \text{Par}_N(k)\}$ or $\{h_\mu : \mu \in \text{Par}_N(k)'\}$	$h_k = \sum_{1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq N} x_{i_1} x_{i_2} \cdots x_{i_k},$ $h_\mu = h_{\mu_1} h_{\mu_2} \cdots h_{\mu_s}$	$\{m_\mu\}$	$\omega(h_\mu) = e_\mu$
Power-sum $\{p_\mu : \mu \in \text{Par}_N(k)'\}$	$p_k = \sum_{i=1}^N x_i^k, p_\mu = \prod_i p_{\mu_i}$	$\{p_\mu/z_\mu\}$	$\omega(p_\mu) = \epsilon_\mu p_\mu$
Schur $\{s_\mu : \mu \in \text{Par}_N(k)\}$	$s_\mu = \sum_{T \in \text{SSYT}_N(\mu)} x^{c(T)}$	$\{s_\mu\}$	$\omega(s_\mu) = s_{\mu'}$

**TABLE 10.3**

Expansions and recursions for symmetric polynomials.

Monomial expansion of Schur polys.:	$s_\lambda = \sum_{\mu \in \text{Par}( \lambda )} K_{\lambda, \mu} m_\mu.$
Schur expansion of complete polys.:	$h_\alpha = \sum_{\lambda \in \text{Par}( \alpha )} K_{\lambda, \alpha} s_\lambda.$
Schur expansion of elementary polys.:	$e_\alpha = \sum_{\lambda \in \text{Par}( \alpha )} K_{\lambda', \alpha} s_\lambda.$
Power-sum expansion of $h_n$ :	$h_n = \sum_{\mu \in \text{Par}(n)} z_\mu^{-1} p_\mu.$
Power-sum expansion of $e_n$ :	$e_n = \sum_{\mu \in \text{Par}(n)} \epsilon_\mu z_\mu^{-1} p_\mu.$
Schur expansion of $p_{(1^n)}$ :	$p_{(1^n)} = \sum_{\lambda \in \text{Par}(n)}  \text{SYT}(\lambda)  s_\lambda.$
Monomial expansion of skew Schur polys.:	$s_{\mu/\nu} = \sum_{\rho \in \text{Par}( \mu/\nu )} K_{\mu/\nu, \rho} m_\rho.$
Schur expansion of $s_\mu h_\alpha$ :	$s_\mu h_\alpha = \sum_{\lambda} K_{\lambda/\mu, \alpha} s_\lambda.$
Schur expansion of $s_\mu e_\alpha$ :	$s_\mu e_\alpha = \sum_{\lambda} K_{\lambda'/\mu', \alpha} s_\lambda.$
Recursion linking $e$ 's and $h$ 's:	$\sum_{i=0}^m (-1)^i e_i h_{m-i} = \chi(m=0).$
Recursion linking $h$ 's and $p$ 's:	$\sum_{i=0}^{n-1} h_i p_{n-i} = n h_n.$
Recursion linking $e$ 's and $p$ 's:	$\sum_{i=0}^{n-1} (-1)^i e_i p_{n-i} = (-1)^{n-1} n e_n.$

of the cells in  $\mu/\nu$  so that rows weakly increase and columns strictly increase. The skew Schur polynomial in  $N$  variables indexed by  $\mu/\nu$  is

$$s_{\mu/\nu}(x_1, \dots, x_N) = \sum_{T \in \text{SSYT}_N(\mu/\nu)} x^{c(T)},$$

where the power of  $x_i$  is the number of  $i$ 's in  $T$ . Skew Schur polynomials are symmetric, since an involution exists that switches the frequencies of  $i$ 's and  $(i+1)$ 's in semistandard tableaux of this shape.

- *Orderings on Partitions.* For  $\mu, \nu \in \text{Par}(k)$ ,  $\mu \leq_{\text{lex}} \nu$  means that  $\mu = \nu$  or the first nonzero entry of  $\nu - \mu$  is positive;  $\leq_{\text{lex}}$  is a total ordering on  $\text{Par}(k)$ . We say  $\mu \leq \nu$  ( $\nu$  dominates  $\mu$ ) iff  $\mu_1 + \dots + \mu_i \leq \nu_1 + \dots + \nu_i$  for all  $i \geq 1$ ;  $\leq$  is a partial ordering on  $\text{Par}(k)$ . We have  $\mu \leq \nu$  iff  $\nu' \leq \mu'$  iff  $\mu$  can be transformed into  $\nu$  by a sequence of raising operators (moving one box to a higher row). Also,  $\mu \leq \nu$  implies  $\mu \leq_{\text{lex}} \nu$ .
- *Kostka Numbers.* For  $\nu \subseteq \mu \in \text{Par}$  and  $\alpha \in \mathbb{N}^N$ , the Kostka number  $K_{\mu/\nu, \alpha}$  is the number of semistandard tableaux of shape  $\mu/\nu$  and content  $\alpha$ . We have  $K_{\lambda, \lambda} = 1$  for all  $\lambda \in \text{Par}$ , and  $K_{\lambda, \mu} \neq 0$  implies  $\mu \leq \lambda$  and  $\mu \leq_{\text{lex}} \lambda$ .
- *Tableau Insertion.* Given a semistandard tableau  $T$  and value  $x$ , we obtain a new semistandard tableau  $T \leftarrow x$  as follows. The element  $x$  bumps the leftmost value  $y > x$  in the top row into the second row, and this bumping continues recursively until a value is placed in a new box at the end of some row. The bumping path moves weakly left as it goes down. Insertion is invertible if we know which corner box is the new one. If we insert a weakly increasing sequence into  $T$ , the new boxes move strictly right and weakly higher, producing a horizontal strip. If we insert a strictly decreasing sequence into  $T$ , the new boxes move weakly left and strictly lower, producing a vertical strip.
- *Pieri Rules.* (a)  $s_\mu h_k = \sum_\nu s_\nu$  where we sum over all  $\nu$  such that  $\nu/\mu$  is a horizontal strip of size  $k$ . (b)  $s_\mu e_k = \sum_\nu s_\nu$  where we sum over all  $\nu$  such that  $\nu/\mu$  is a vertical strip of size  $k$ . If there are  $N$  variables, we only use shapes  $\nu$  with at most  $N$  parts.
- *Algebraic Independence.* A list of polynomials  $(f_1, \dots, f_k)$  is algebraically independent over  $K$  iff the only polynomial  $g \in K[z_1, \dots, z_k]$  with  $g(f_1, \dots, f_k) = 0$  is the zero polynomial. Equivalently, the monomials  $\{f_1^{i_1} \dots f_k^{i_k} : i_j \in \mathbb{N}\}$  are linearly independent over  $K$ . Polynomials  $f_1, \dots, f_k \in K[x_1, \dots, x_k]$  are algebraically independent over  $K$  if (and only if)  $\det \|D_i f_j\|_{1 \leq i, j \leq k} \neq 0$ .
- *Algebraically Independent Symmetric Polynomials.* In the ring  $K[x_1, \dots, x_N]$ , the lists  $(p_1, \dots, p_N)$ ,  $(h_1, \dots, h_N)$ , and  $(e_1, \dots, e_N)$  are algebraically independent over  $K$ . So we can view the ring and  $K$ -vector space  $\Lambda_N$  as isomorphic to the polynomial ring  $K[z_1, \dots, z_N]$  in three ways (an isomorphism can send  $z_i \in K[z_1, \dots, z_N]$  to either  $p_i$  or  $h_i$  or  $e_i$ ).
- *Generating Functions for  $e$ 's and  $h$ 's.* We have

$$E_N(t) = \prod_{i=1}^N (1 + x_i t) = \sum_{k=0}^N e_k(x_1, \dots, x_N) t^k;$$

$$H_N(t) = \prod_{i=1}^N (1 - x_i t)^{-1} = \sum_{k=0}^N h_k(x_1, \dots, x_N) t^k;$$

so  $H_N(t)E_N(-t) = 1$ .

- *Dual Bases and Cauchy Identities.* Assume  $N \geq k$ . The Hall scalar product on  $\Lambda_N^k$  is defined by setting  $\langle p_\mu, p_\nu \rangle = z_\mu \chi(\mu = \nu)$  and extending by bilinearity. Two bases  $\{f_\mu : \mu \in \text{Par}(k)\}$  and  $\{g_\mu : \mu \in \text{Par}(k)\}$  of  $\Lambda_N^k$  are dual relative to this inner product iff they satisfy the Cauchy identity

$$\text{coefficient of } t^k \text{ in } \prod_{i=1}^N \prod_{j=1}^N \frac{1}{1 - x_i y_j t} = \sum_{\mu \in \text{Par}(k)} f_\mu(x_1, \dots, x_N) g_\mu(y_1, \dots, y_N).$$

- *The Map  $\omega$ .* The map  $\omega : \Lambda_N \rightarrow \Lambda_N$  is defined by sending  $c$  to  $c$  (for  $c \in K$ ) and  $p_i$  to  $(-1)^{i-1} p_i$ , and extending by the universal mapping property of polynomial rings. Note  $\omega(p_\mu) = \epsilon_\mu p_\mu$  where  $\epsilon_\mu = (-1)^{|\mu| - \ell(\mu)}$ . The map  $\omega$  is an involution ( $\omega^2 = \text{id}$ ), an isomorphism of rings and vector spaces, and (for all  $k \leq N$ ) an isometry of  $\Lambda_N^k$  (so  $\langle \omega(f), \omega(g) \rangle = \langle f, g \rangle$  for  $f, g \in \Lambda_N^k$ ).
- *RSK Correspondences.* There are bijections between: (a) permutations in  $S_n$  and pairs  $(P, Q)$  of standard tableaux of the same shape  $\lambda \in \text{Par}(n)$ ; (b) words in  $X^n$  and pairs  $(P, Q)$  where  $P \in \text{SSYT}_X(\lambda)$  and  $Q \in \text{SYT}(\lambda)$  for some  $\lambda \in \text{Par}(n)$ ; (c)  $M \times N$  matrices with values in  $\mathbb{N}$  and pairs  $(P, Q)$  where  $P \in \text{SSYT}_N(\lambda)$  and  $Q \in \text{SSYT}_M(\lambda)$ . In each case, one inserts successive entries into  $P$ , using  $Q$  to record the locations of new boxes. For (c), one must first encode the matrix as a biword. If  $w \in S_n$  maps to  $(P, Q)$ , then  $w^{-1}$  maps to  $(Q, P)$ . If  $w \in X^n$  maps to  $(P, Q)$ , then  $\text{Des}(w) = \text{Des}(Q)$ ,  $\text{des}(w) = \text{des}(Q)$ , and  $\text{maj}(w) = \text{maj}(Q)$ , where  $\text{Des}(Q)$  is the set of  $k < n$  such that  $k+1$  is in a lower row of  $Q$  than  $k$ ,  $\text{des}(Q) = |\text{Des}(Q)|$ , and  $\text{maj}(Q) = \sum_{i \in \text{Des}(Q)} i$ .

## Exercises

- 10.135.** Draw all skew shapes  $\mu/\nu$  where  $\mu \vdash 6$  and  $\nu \vdash 3$ . Indicate which skew shapes are horizontal (resp. vertical) strips.
- 10.136.** Given a skew shape  $S \subseteq \mathbb{N} \times \mathbb{N}$ , describe how to calculate the number of different pairs of partitions  $(\mu, \nu)$  such that  $S = \mu/\nu$ .
- 10.137.** Find necessary and sufficient algebraic conditions on the parts of  $\mu$  and  $\nu$  to ensure that the skew shape  $\mu/\nu$  is (a) a horizontal strip; (b) a vertical strip.
- 10.138.** How many horizontal strips are contained in  $\{1, 2, \dots, a\} \times \{1, 2, \dots, b\}$ ?
- 10.139.** If  $|\mu/\nu| = n$  and  $|X| = k$ , how many tableaux with values in  $X$  have shape  $\mu/\nu$ ?
- 10.140.** List all the tableaux in: (a)  $\text{SSYT}_5((3, 2))$ ; (b)  $\text{SSYT}_2((3, 2))$ ; (c)  $\text{SYT}((3, 2, 1))$ .
- 10.141.** Give a direct counting argument to determine  $|\text{SYT}(\mu)|$  when  $\mu$  is a hook.
- 10.142.** Prove that  $|\text{SYT}(\mu/\nu)| = |\text{SYT}(\mu'/\nu')|$  for all skew shapes  $\mu/\nu$ .
- 10.143.** Compute  $s_{(2,2)}(x_1, \dots, x_N)$  for  $N = 3, 4, 5$  by enumerating tableaux.
- 10.144.** Compute  $s_{(2,2)/(1)}(x_1, \dots, x_N)$  for  $N = 2, 3, 4$  by enumerating tableaux.
- 10.145.** Find the coefficients of the following monomials in  $s_{(3,2,1)}(x_1, \dots, x_6)$  by enumerating tableaux: (a)  $x_1 x_2 x_3 x_4 x_5 x_6$ ; (b)  $x_1^2 x_2^2 x_3^2$ ; (c)  $x_1^3 x_2^3$ ; (d)  $x_1^2 x_2 x_3 x_4 x_5$ ; (e)  $x_1 x_2 x_3^2 x_4 x_5$ ; (f)  $x_1 x_2 x_3 x_4 x_5^2$ .

**10.146.** Let  $N \geq 4$ . Enumerate tableaux to confirm that the coefficients of  $x_1^2 x_2 x_3^2 x_4$ ,  $x_1 x_2^2 x_3 x_4^2$ , and  $x_1 x_2^2 x_3^2 x_4$  in  $s_{(4,3)/(1)}(x_1, \dots, x_N)$  are all equal to 6, as claimed in 10.17. What happens to these coefficients if  $N < 4$ ?

**10.147.** For which values of  $N$  is  $s_{\mu/\nu}(x_1, \dots, x_N) = 0$ ?

**10.148.** Compute: (a)  $p_4(x_1, x_2, x_3)$ ; (b)  $e_3(x_1, \dots, x_5)$ ; (c)  $h_3(x_1, x_2, x_3)$ ; (d)  $m_{(3,2,2)}(x_1, \dots, x_4)$ .

**10.149.** For  $\mu = (2, 1)$ , compute: (a)  $p_\mu(x_1, x_2, x_3)$ ; (b)  $e_\mu(x_1, x_2, x_3)$ ; (c)  $h_\mu(x_1, x_2, x_3)$ .

**10.150.** How many monomials appear with nonzero coefficient in: (a)  $e_k(x_1, \dots, x_n)$ ; (b)  $h_k(x_1, \dots, x_n)$ ?

**10.151.** How many monomials appear with nonzero coefficient in  $m_\mu(x_1, \dots, x_N)$ ?

**10.152.** Give a direct proof that the polynomials  $e_k(x_1, \dots, x_N)$  and  $h_k(x_1, \dots, x_N)$  (as defined in 10.21 and 10.22) are symmetric.

**10.153.** Find a skew shape  $\mu/\nu$  such that  $e_3 h_4 h_2 e_5 h_1 = s_{\mu/\nu}$ .

**10.154.** Prove that any finite product of skew Schur polynomials is a skew Schur polynomial.

**10.155.** Check that the set of homogeneous polynomials of degree  $k$  in  $R = K[x_1, \dots, x_N]$  is a vector subspace of  $R$ . Conclude that  $\Lambda_N^k$  is a subspace of  $\Lambda_N$  for each  $k \geq 0$ .

**10.156.** Write down an explicit basis for the  $K$ -vector space  $\Lambda_3^7$ .

**10.157.** Compute  $\dim(\Lambda_N^k)$  for each choice of  $k$  and  $N$  in the range  $1 \leq k \leq 6$  and  $1 \leq N \leq 6$ .

**10.158.** Suppose  $\{f_i : i \in I\}$  is a collection of nonzero polynomials in  $R = K[x_1, \dots, x_N]$  such that, whenever some  $x^\alpha$  appears in some  $f_i$  with nonzero coefficient, the coefficient of  $x^\alpha$  in every other  $f_j$  is zero. Prove that  $\{f_i : i \in I\}$  is linearly independent over  $K$ .

**10.159.** Compute the following Kostka numbers:

(a)  $K_{(3,3,2),(2,1,2,1,1,1)}$ ; (b)  $K_{(3,2,2,1),(2,2,1,1,1,1)}$ ; (c)  $K_{(5,5),(1^{10})}$ ; (d)  $K_{(3,3,3)/(2,1),(2,2,1,1)}$ .

**10.160.** Compute the image of the first tableau in the proof of 10.33 under the maps  $f_i$ , for  $i = 1, 2, 4, 5, 6, 7, 8$ .

**10.161.** Use the maps  $f_i$  in the proof of 10.33 to compute specific bijections between the three collections of six tableaux in 10.17. (This calculation was begun in 10.34.)

**10.162.** Express the Schur polynomials  $s_\mu(x_1, x_2, x_3, x_4, x_5)$  as explicit linear combinations of monomial symmetric polynomials, for all  $\mu \vdash 4$  and  $\mu \vdash 5$ .

**10.163.** Express the skew Schur polynomial  $s_{(3,3,2)/(1)}(x_1, \dots, x_8)$  as a linear combination of monomial symmetric polynomials.

**10.164.** For all partitions  $\mu \vdash 3$ , express  $h_\mu$  and  $e_\mu$  in terms of monomial symmetric polynomials by viewing  $h_\mu$  and  $e_\mu$  as instances of skew Schur polynomials.

**10.165.** (a) Find a recursion characterizing the Kostka numbers  $K_{\mu/\nu, \alpha}$ . (b) Use (a) to write a computer program for computing Kostka numbers.

**10.166.** Check that  $\leq_{\text{lex}}$  is a total ordering of the set  $\text{Par}(k)$ , for each  $k \geq 0$ .

**10.167.** Prove that  $\leq$  is a total ordering of  $\text{Par}(k)$  iff  $k \leq 5$ .

**10.168.** (a) List the integer partitions of 7 in lexicographic order. (b) Find all pairs  $\mu, \nu \vdash 7$  such that  $\mu \leq_{\text{lex}} \nu$  but  $\mu \not\leq \nu$ .

**10.169.** (a) Find an ordered sequence of raising operators that changes  $\mu = (5, 4, 2, 1, 1)$  to  $\nu = (7, 3, 2, 1)$ . (b) How many such sequences are there?

**10.170.** Prove or disprove: for all partitions  $\mu, \nu \vdash k$ ,  $\mu \leq_{\text{lex}} \nu$  iff  $\nu' \leq_{\text{lex}} \mu'$ .

**10.171.** Let  $\mu, \nu \vdash k$ . Can you prove that  $\mu \leq \nu$  implies  $\nu' \leq \mu'$  by arguing directly from the definitions, without using raising operators?

**10.172.** Define an ordering  $\leq_{\text{lex}}$  on the set  $\mathbb{N}^N$  as in 10.36. Show that  $\leq_{\text{lex}}$  is a total ordering of  $\mathbb{N}^N$  satisfying the following *well-ordering* property: there exists no infinite strictly decreasing sequence

$$\alpha^{(1)} >_{\text{lex}} \alpha^{(2)} >_{\text{lex}} \cdots >_{\text{lex}} \alpha^{(k)} >_{\text{lex}} \cdots \quad (\alpha^{(j)} \in \mathbb{N}^N).$$

**10.173.** Define the *lex degree* of a nonzero polynomial  $f(x_1, \dots, x_N) \in R = K[x_1, \dots, x_N]$ , denoted  $\deg(f)$ , to be the largest  $\alpha \in \mathbb{N}^N$  (relative to the lexicographic ordering defined in 10.172) such that  $x^\alpha$  occurs with nonzero coefficient in  $f$ . Prove that  $\deg(gh) = \deg(g) + \deg(h)$  for all nonzero  $g, h \in R$ , and  $\deg(g+h) \leq \max(\deg(g), \deg(h))$  whenever both sides are defined.

**10.174.** (a) Find the Kostka matrix indexed by all partitions of 4. (b) Invert this matrix, and thereby express the monomial symmetric polynomials  $m_\mu(x_1, x_2, x_3, x_4)$  (for  $\mu \vdash 4$ ) as linear combinations of Schur polynomials.

**10.175.** Find the Kostka matrix indexed by partitions in  $\text{Par}_3(7)$ , and invert it.

**10.176.** Let  $\mathbf{K}$  be the Kostka matrix indexed by all partitions of 8. How many nonzero entries does this matrix have?

**10.177.** Suppose  $A$  is an  $n \times n$  matrix with integer entries such that  $\det(A) = \pm 1$ . Prove that  $A^{-1}$  has all integer entries. (In particular, this applies when  $A$  is a Kostka matrix.)

**10.178.** Suppose  $\{v_i : i \in I\}$  is a basis for a finite-dimensional  $K$ -vector space  $V$ ,  $\{w_i : i \in I\}$  is an indexed family of vectors in  $V$ , and for some total ordering  $\leq$  of  $I$  and some scalars  $a_{ij} \in K$  with  $a_{ii} \neq 0$ , we have  $w_i = \sum_{j \leq i} a_{ij} v_j$  for all  $i \in I$ . Prove that  $\{w_i : i \in I\}$  is a basis of  $V$ .

**10.179.** Let  $T$  be the tableau in 10.53. Confirm that  $T \leftarrow 1$  and  $T \leftarrow 0$  are as stated in that example. Also, compute  $T \leftarrow i$  for  $i = 2, 4, 5, 7$ , and verify that 10.54 holds.

**10.180.** Let  $T$  be the semistandard tableau

2	2	3	5	5	7	7
3	3	4	6	7	8	
4	5	5	8	8		
6	6	6	9			
7	8	8				
8						

Compute  $T \leftarrow i$  for  $1 \leq i \leq 9$ .

**10.181.** Suppose we apply the tableau insertion algorithm 10.52 to a tableau  $T$  of skew shape. Are 10.54 and 10.55 still true?

**10.182.** Give a non-recursive description of  $T \leftarrow x$  in the case where: (a)  $x$  is larger than every entry of  $T$ ; (b)  $x$  is smaller than every entry of  $T$ .

**10.183.** Let  $T$  be the tableau in 10.53. Perform reverse insertion starting at each corner box of  $T$  to obtain smaller tableaux  $T_i$  and values  $x_i$ . Verify that  $T_i \leftarrow x_i = T$  for each answer.

**10.184.** Let  $T$  be the tableau in 10.180. Perform reverse insertion starting at each corner box of  $T$ , and verify that properties 10.58(a),(b) hold in each case.

**10.185.** Prove 10.58(c).

**10.186.** Prove 10.59.

**10.187.** Express  $s_{(4,4,3,1,1)}h_1$  as a sum of Schur polynomials.

**10.188.** Let  $T$  be the tableau in 10.53. Successively insert 1, 2, 2, 3, 5, 5 into  $T$ , and verify that the assertions of the bumping comparison theorem hold.

**10.189.** Let  $T$  be the tableau in 10.53. Successively insert 7, 5, 3, 2, 1 into  $T$ , and verify that the assertions of the bumping comparison theorem hold.

**10.190.** Let  $T$  be the tableau in 10.180. Successively insert 1, 1, 3, 3, 3, 4 into  $T$ , and verify that the assertions of the bumping comparison theorem hold.

**10.191.** Let  $T$  be the tableau in 10.180. Successively insert 7, 6, 5, 3, 2, 1 into  $T$ , and verify that the assertions of the bumping comparison theorem hold.

**10.192.** Let  $T$  be the tableau in 10.61 of shape  $\mu = (5, 4, 4, 4, 1)$ . For each shape  $\nu$  such that  $\nu/\mu$  is a horizontal strip of size 3, find a weakly increasing sequence  $x_1 \leq x_2 \leq x_3$  such that  $((T \leftarrow x_1) \leftarrow x_2) \leftarrow x_3$  has shape  $\nu$ , or prove that no such sequence exists.

**10.193.** Repeat the previous exercise, replacing horizontal strips by vertical strips and weakly increasing sequences by strictly decreasing sequences.

**10.194.** Prove 10.64(b).

**10.195.** Let  $T$  be the tableau in 10.180. Find the unique semistandard tableau  $S$  of shape  $(7, 5, 4, 4, 1, 1)$  and  $z_1 \leq z_2 \leq z_3 \leq z_4$  such that  $T = S \leftarrow z_1 z_2 z_3 z_4$ .

**10.196.** Let  $T$  be the tableau in 10.180. Find the unique semistandard tableau  $S$  of shape  $(6, 6, 5, 3, 2)$  and  $z_1 > z_2 > z_3 > z_4$  such that  $T = S \leftarrow z_1 z_2 z_3 z_4$ .

**10.197.** Expand each symmetric polynomial into sums of Schur polynomials: (a)  $s_{(4,3,1)}e_2$ ; (b)  $s_{(2,2)}h_3$ ; (c)  $s_{(2,2,1,1,1)}h_4$ ; (d)  $s_{(3,3,2)}e_3$ .

**10.198.** Use the Pieri rule to find the Schur expansions of  $h_{(3,2,1)}$ ,  $h_{(3,1,2)}$ ,  $h_{(1,2,3)}$ , and  $h_{(1,3,2)}$ , and verify that the answers agree with those found in 10.67.

**10.199.** Expand each symmetric polynomial into sums of Schur polynomials: (a)  $h_{(2,2,2)}$ ; (b)  $h_{(5,3)}$ ; (c)  $s_{(3,2)}h_{(2,1)}$ ; (d)  $s_{(6,3,2,2)}/(3,2)$ .

**10.200.** Find the coefficients of the following Schur polynomials in the Schur expansion of  $h_{(3,2,2,1,1)}$ : (a)  $s_{(9)}$ ; (b)  $s_{(5,4)}$ ; (c)  $s_{(4,4,1)}$ ; (d)  $s_{(2,2,2,2,1)}$ ; (e)  $s_{(3,3,3)}$ ; (f)  $s_{(3,2,2,1,1)}$ .

**10.201.** Use 10.73 to compute the monomial expansions of  $h_\mu(x_1, x_2, x_3, x_4)$  for all partitions  $\mu$  of size at most four.

**10.202.** Let  $\alpha = (\alpha_1, \dots, \alpha_s)$ . Prove that the coefficient of  $m_\lambda(x_1, \dots, x_N)$  in the monomial expansion of  $h_\alpha(x_1, \dots, x_N)$  is the number of  $s \times N$  matrices  $A$  with entries in  $\mathbb{N}$  such that  $\sum_{j=1}^N A(i, j) = \alpha_i$  for  $1 \leq i \leq s$  and  $\sum_{i=1}^s A(i, j) = \lambda_j$  for  $1 \leq j \leq N$ .

**10.203.** Use the Pieri rules to compute the Schur expansions of: (a)  $e_{(3,3,1)}$ ; (b)  $e_{(5,3)}$ ; (c)  $s_{(3,2)}e_{(2,1)}$ ; (d)  $s_{(4,3,3,3,1,1)}/(3,1,1,1)$ .

**10.204.** Find the coefficients of the following Schur polynomials in the Schur expansion of  $e_{(4,3,2,1)}$ : (a)  $s_{(4,3,2,1)}$ ; (b)  $s_{(5,5)}$ ; (c)  $s_{(2,2,2,2,2)}$ ; (d)  $s_{(2,2,2,1^4)}$ ; (e)  $s_{(1^{10})}$ .

**10.205.** Use 10.78 to express  $e_{(2,2,1)}$  and  $e_{(3,2)}$  as linear combinations of monomial symmetric polynomials.

**10.206.** Prove the formula for  $s_\mu e_\alpha$  stated in Table 10.3.

**10.207.** Let  $\alpha = (\alpha_1, \dots, \alpha_s)$ . Find a combinatorial interpretation for the coefficient of  $m_\lambda(x_1, \dots, x_N)$  in the monomial expansion of  $e_\alpha(x_1, \dots, x_N)$  in terms of certain  $s \times N$  matrices (cf. 10.202).

**10.208.** Prove that the following lists of polynomials are algebraically dependent by exhibiting an explicit dependence relation: (a)  $h_i(x_1, x_2)$  for  $1 \leq i \leq 3$ ; (b)  $e_i(x_1, x_2, x_3)$  for  $1 \leq i \leq 4$ ; (c)  $p_i(x_1, x_2, x_3)$  for  $1 \leq i \leq 4$ .

**10.209.** Prove that any sublist of an algebraically independent list is algebraically independent.

**10.210.** Suppose  $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{N}^N$  is a partition. Show that  $\deg(e_1^{\alpha_1 - \alpha_2} e_2^{\alpha_2 - \alpha_3} \dots e_{N-1}^{\alpha_{N-1} - \alpha_N} e_N^{\alpha_N}) = \alpha$  (see 10.173 for the definition of lex degree).

**10.211. Algorithmic Proof of the Fundamental Theorem of Symmetric Polynomials.** Prove that the following algorithm will express any  $f \in \Lambda_N$  as a polynomial in the elementary symmetric polynomials  $e_i(x_1, \dots, x_N)$  (where  $1 \leq i \leq N$ ) in finitely many steps. If  $f = 0$ , use the zero polynomial. Otherwise, let the term of largest degree in  $f$  (see 10.173) be  $cx^\alpha$  where  $c \in K$  is nonzero. Use symmetry of  $f$  to show that  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_N$ , and that  $f - ce_1^{\alpha_1 - \alpha_2} e_2^{\alpha_2 - \alpha_3} \dots e_{N-1}^{\alpha_{N-1} - \alpha_N} e_N^{\alpha_N}$  is either 0 or has degree  $\beta <_{\text{lex}} \alpha$  (see 10.210). Continue similarly to express this new polynomial (and hence  $f$ ) as a polynomial in the  $e_i$ 's.

**10.212.** Use the algorithm in the preceding exercise to express  $m_{(2,1)}(x_1, x_2, x_3, x_4)$  and  $p_3(x_1, x_2, x_3, x_4)$  as a polynomial in  $\{e_i(x_1, x_2, x_3, x_4) : 1 \leq i \leq 4\}$ .

**10.213.** Use the test in 10.83 to verify that the polynomials  $h_i(x_1, x_2, x_3)$  for  $1 \leq i \leq 3$  are algebraically independent. Can you generalize this computation to more than 3 variables?

**10.214.** Use the test in 10.83 to verify that the polynomials  $e_i(x_1, x_2, x_3, x_4)$  for  $1 \leq i \leq 4$  are algebraically independent. Can you generalize this computation to more than 4 variables?

**10.215.** Compute the images of

$$\left( 4, \begin{array}{|c|} \hline 1 \\ \hline 3 \\ \hline 4 \\ \hline 5 \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline 1 & 2 & 2 & 3 \\ \hline \end{array} \right) \text{ and } \left( 4, \begin{array}{|c|} \hline 2 \\ \hline 3 \\ \hline 4 \\ \hline 5 \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & 3 \\ \hline \end{array} \right)$$

under the involution  $I$  in the proof of 10.87.

**10.216.** Write out all the matched pairs  $(z, I(z))$  in the proof of 10.87 when: (a)  $N = 2$  and  $m = 3$ ; (b)  $N = 3$  and  $m = 2$ .

**10.217.** Imitate the argument in the proof of 10.88 to show that algebraic independence of  $(h_1, \dots, h_N)$  in  $K[x_1, \dots, x_N]$  implies algebraic independence of  $(e_1, \dots, e_N)$ .

**10.218.** (a) Prove the recursion  $e_k(x_1, \dots, x_N) = e_k(x_1, \dots, x_{N-1}) + e_{k-1}(x_1, \dots, x_{N-1})x_N$  for  $k, N \geq 1$ . What are the initial conditions? (b) Find a similar recursion for  $h_k(x_1, \dots, x_N)$ .

**10.219.** (a) Prove  $s'(n, k) = e_{n-k}(1, 2, \dots, n-1)$ . (b) Prove  $S(n, k) = h_{n-k}(1, 2, \dots, k)$ .

**10.220.** Prove 10.91 by expanding  $\prod_{i=1}^N (X - r_i)$  using the generalized distributive law.

**10.221.** Consider the polynomial  $p = x^5 - 2x^4 + 5x^3 + 7x^2 - x - 4$ , which has five roots  $r_1, \dots, r_5 \in \mathbb{C}$ . Compute: (a) the sum of the roots; (b) the product of the roots; (c)  $e_3(r_1, \dots, r_5)$ ; (d) the sum of the squares of the roots; (e)  $\sum_{i \neq j} r_i^2 r_j$ .

**10.222.** Use 10.92 to calculate the coefficient of  $t^4$  in the multiplicative inverse of  $(1 - 2x)(1 - 3x)(1 - 5x)$ .

**10.223.** Let  $A$  be an  $n \times n$  complex matrix. What is the relationship between the coefficients of the characteristic polynomial  $\det(tI - A)$  and the eigenvalues  $r_1, \dots, r_n$  of  $A$ ?

**10.224.** Use (10.8) to show that  $p_i(x_1, \dots, x_N)$  (for  $1 \leq i \leq N$ ) are algebraically independent over  $K$  iff  $h_i(x_1, \dots, x_N)$  (for  $1 \leq i \leq N$ ) are algebraically independent over  $K$ .

**10.225.** Use (10.9) to show that  $p_i(x_1, \dots, x_N)$  (for  $1 \leq i \leq N$ ) are algebraically independent iff  $e_i(x_1, \dots, x_N)$  (for  $1 \leq i \leq N$ ) are algebraically independent.

**10.226.** Consider the maps  $f$  and  $g$  from the proof of 10.93. Compute

$$f((5, \boxed{2 \mid 4 \mid 4 \mid 5 \mid 5}, \boxed{4 \mid 4})) \text{ and } g(\boxed{1 \mid 1^* \mid 1 \mid 2 \mid 2 \mid 4 \mid 6 \mid 6}).$$

**10.227.** Consider the maps  $I$  and  $g$  from the proof of 10.94. Compute

$$I\left(3, \boxed{\frac{1}{3}}, \boxed{3 \mid 3 \mid 3}\right), \quad I\left(3, \boxed{\frac{1}{3}}, \boxed{2 \mid 2 \mid 2}\right), \quad I\left(3, \boxed{\frac{1}{3}}, \boxed{5}\right), \quad I\left(3, \boxed{\frac{1}{3}}, \boxed{4}\right).$$

For any objects that are fixed points of  $I$ , compute the images of those objects under  $g$ .

**10.228.** Write  $\sum_{i=1}^N \frac{x_i}{1-x_i t}$  in terms of suitable symmetric polynomials.

**10.229.** Obtain 10.93 and 10.94 algebraically by differentiating the generating functions  $H_N(t)$  and  $E_N(-t)$ .

**10.230.** Use the recursions 10.93 and 10.94 to verify the formulas for  $h_4$ ,  $3!e_3$ , and  $4!e_4$  stated in 10.95.

**10.231.** Complete the proof of 10.96 by checking that  $g(f(y_0)) = y_0$  and  $f(g(z_0)) = z_0$ , and, in general,  $g \circ f = \text{id}_Y$  and  $f \circ g = \text{id}_X$ .

**10.232.** Let  $g$  be the map in the proof of 10.96. Compute  $g(z_1)$  and  $g(z_2)$ , where

$$z_1 = \begin{bmatrix} w : & 3 & 7 & 2 & 5 & 4 & 6 & 8 & 1 \\ T : & 2 & 4 & 4 & 4 & 4 & 6 & 6 & 6 \end{bmatrix}; \quad z_2 = \begin{bmatrix} w : & 2 & 1 & 4 & 3 & 7 & 5 & 6 & 8 \\ T : & 1 & 1 & 1 & 2 & 3 & 3 & 3 & 3 \end{bmatrix}.$$



**10.233.** Let  $f$  be the map in the proof of 10.96. Compute  $f(y)$ , where

$$y = \left( (2, 2, 2, 2), (2, 5)(3, 8)(4, 6)(1, 7), \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 2 & 3 & 2 & 2 & 2 & 3 & 3 \end{pmatrix} \right).$$

**10.234.** Let  $I$  be the involution in the proof of 10.98. Compute  $I(z_1)$ ,  $I(z_2)$ , and  $I(f(y))$ , where  $z_1$ ,  $z_2$ , and  $y$  are the objects given in the preceding two exercises.

**10.235.** Let  $A$  be an  $n \times n$  complex matrix with eigenvalues  $r_1, \dots, r_n$ . (a) Show that the trace of  $A$ , defined by  $\text{tr}(A) = \sum_{i=1}^n A(i, i)$ , is  $p_1(r_1, \dots, r_n)$ . (b) For  $k \geq 1$ , express  $\text{tr}(A^k)$  as a function of  $r_1, \dots, r_n$ . (c) Suppose  $n = 5$  and  $(\text{tr}(A^k) : k = 1, 2, \dots, 5) = (3, 41, -93, 693, -2957)$ . Find the characteristic polynomial of  $A$ .

**10.236.** Compute: (a)  $\omega(h_3)$ ; (b)  $\omega(p_{(3,2,1,1)})$ ; (c)  $\omega(e_{(4,4)})$ ; (d)  $\omega(s_{(5,3,3,1,1,1)})$ .

**10.237.** Show that there exists a unique automorphism of the ring and  $K$ -vector space  $\Lambda_N$  mapping each  $c \in K$  to itself and sending each  $p_i$  to  $-p_i$ . Compute the image of  $h_n$  and of  $e_n$  under this automorphism.

**10.238.** In the proof of 10.101(b), where is the assumption  $n \leq N$  needed?

**10.239.** Compute the polynomials  $\text{fgt}_\lambda(x_1, x_2, x_3)$  for all partitions of size at most 3.

**10.240.** Compute  $\text{RSK}(w)$  for all  $w \in S_3$ .

**10.241.** Compute  $\text{RSK}^{-1}(P, Q)$  for all pairs  $P, Q$  of standard tableaux of shape  $(2, 2)$ .

**10.242.** Let  $w = 41572863 \in S_8$ . Compute  $\text{RSK}(w)$  and  $\text{RSK}(w^{-1})$ . Verify that 10.112 holds in this case.

**10.243.** Consider the pair of standard tableaux

$$P = \begin{array}{|c|c|c|} \hline 1 & 3 & 6 \\ \hline 2 & 4 & 8 \\ \hline 5 & 7 & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|} \hline 1 & 2 & 5 \\ \hline 3 & 6 & 7 \\ \hline 4 & 8 & \\ \hline \end{array}.$$

Compute  $w = \text{RSK}^{-1}(P, Q)$  and  $v = \text{RSK}^{-1}(Q, P)$ , and verify that 10.112 holds in this case.

**10.244.** (a) Verify that 10.109 holds for the example  $w = 35164872$  by comparing the first rows of the tableaux in Figure 10.1 with the shadow diagram in Figure 10.4. (b) Similarly, verify the assertions in 10.111 using Figure 10.5.

**10.245.** Draw all the shadow diagrams for the permutations  $w$  and  $w^{-1}$  in 10.242, and use them to verify the assertions in 10.111 for this example.

**10.246.** Draw all the shadow diagrams for the permutations  $w$  and  $v$  in 10.243, and use them to verify the assertions in 10.111 for this example.

**10.247.** (a) Point out why  $n! = \sum_{\lambda \vdash n} |\text{SYT}(\lambda)|^2$ . (b) Verify this identity directly for  $n = 5$ .

**10.248.** Show that the number of  $w \in S_n$  such that  $w^2 = \text{id}$  is given by  $\sum_{\lambda \vdash n} |\text{SYT}(\lambda)|$ .

**10.249.** Compute  $\text{RSK}(w)$  for all words  $w \in \{0, 1\}^3$ .

**10.250.** Compute  $\text{RSK}(313211231)$ , and verify that 10.117 holds in this case.

**10.251.** Compute the word  $w$  such that

$$\text{RSK}(w) = \left( \begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & 2 \\ \hline 2 & 3 & 4 & 4 \\ \hline 4 & 5 & 5 & \\ \hline \end{array}, \begin{array}{|c|c|c|c|} \hline 1 & 2 & 4 & 6 \\ \hline 3 & 5 & 8 & 10 \\ \hline 7 & 9 & 11 & \\ \hline \end{array} \right).$$

Verify that 10.117 holds.

**10.252.** (a) Compute  $\sum_{T \in \text{SYT}((4,1))} q^{\text{maj}(T)}$ . (b) Compute  $\sum_{T \in \text{SYT}((3,2,1))} q^{\text{maj}(T)}$ .

**10.253.** Express  $p_{(1^4)}$  as a linear combination of Schur polynomials.

**10.254.** (a) Compute the biword and the pair of tableaux associated to the matrix  $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 3 & 2 & 0 \end{bmatrix}$ . (b) Do the same for the transpose of this matrix.

**10.255.** (a) Compute the matrix and pair of tableaux associated to the biword

$$\left( \begin{array}{ccccccccc} 1 & 1 & 2 & 2 & 2 & 3 & 3 & 5 \\ 2 & 4 & 1 & 1 & 3 & 3 & 3 & 2 \end{array} \right).$$

(b) Do the same for the biword obtained by switching the two rows and sorting the new top row into increasing order (using the values in the bottom row to break ties).

**10.256.** (a) Compute the biword and matrix associated to the pair of tableaux

$$P = \begin{array}{|c|c|c|} \hline 1 & 1 & 3 \\ \hline 2 & 3 & \\ \hline 4 & 4 & \\ \hline 5 & 5 & \\ \hline \end{array}, \quad Q = \begin{array}{|c|c|c|} \hline 1 & 2 & 2 \\ \hline 3 & 3 & \\ \hline 4 & 5 & \\ \hline 6 & 6 & \\ \hline \end{array}.$$

(b) Do the same for the pair of tableaux  $(Q, P)$ .

**10.257.** Show that if a matrix  $A$  maps to  $(P, Q)$  under the RSK correspondence, then the transposed matrix  $A^t$  maps to  $(Q, P)$  under RSK. Do this by generalizing the shadow constructions in §10.22, allowing more than one dot to occupy a given point  $(i, j)$  in the graph.

**10.258.** Give a rigorous justification of the computation

$$\prod_{j=1}^N \sum_{k_j=0}^{\infty} h_{k_j}(x_1, \dots, x_M) y_j^{k_j} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \prod_{j=1}^N h_{k_j}(x_1, \dots, x_M) y_j^{k_j},$$

which was used in the proof of 10.128.

**10.259.** Verify the fact (used in the proof of 10.131) that the polynomials

$$\{p_{\alpha}(\vec{x})p_{\beta}(\vec{y})/z_{\beta} : (\alpha, \beta) \in \text{Par}(k) \times \text{Par}(k)\}$$

are linearly independent.

**10.260.** Suppose  $A$  and  $B$  are  $n \times n$  matrices with entries in  $K$  such that  $AB = I$ . Prove that  $BA = I$ .

**10.261.** Suppose  $\{f_{\mu} : \mu \in \text{Par}(k)\}$  is an orthonormal basis of  $\Lambda_N^k$ , and  $g \in \Lambda_N^k$ . Prove that

$$g = \sum_{\mu \in \text{Par}(k)} \langle g, f_{\mu} \rangle f_{\mu}.$$

**10.262. Quasisymmetric Polynomials.** A polynomial  $f \in K[x_1, \dots, x_N]$  is called *quasisymmetric* iff for all compositions  $\alpha = (\alpha_1, \dots, \alpha_s)$  with  $s \leq N$  and all  $1 \leq i_1 < i_2 < \dots < i_s \leq N$ , the coefficient of  $\prod_{j=1}^s x_{i_j}^{\alpha_j}$  in  $f$  equals the coefficient of  $\prod_{j=1}^s x_j^{\alpha_j}$  in  $f$ . For each such  $\alpha$ , define the *monomial quasisymmetric polynomial*  $M_\alpha = \sum_{1 \leq i_1 < i_2 < \dots < i_s \leq N} \prod_{j=1}^s x_{i_j}^{\alpha_j}$ . (a) Show that the quasisymmetric polynomials form a subspace of  $K[x_1, \dots, x_N]$  with basis  $\{M_\alpha\}$ . (b) What is the dimension of the space of homogeneous quasisymmetric polynomials of degree  $k$ ? (c) Show that every symmetric polynomial is quasisymmetric. More specifically, express each  $m_\lambda$  in terms of the  $M_\alpha$ 's.

**10.263. Fundamental Quasisymmetric Polynomials.** For each  $n \geq 0$  and each subset  $S$  of  $\{1, 2, \dots, n-1\}$ , define  $Q_{n,S}(x_1, \dots, x_N) = \sum x_{i_1} x_{i_2} \cdots x_{i_n}$  where we sum over all sequences  $1 \leq i_1 \leq i_2 \leq \dots \leq i_n \leq N$  such that  $j \in S$  implies  $i_j < i_{j+1}$ .  $Q_{n,S}$  is called a *fundamental quasisymmetric polynomial*. (a) Show that  $Q_{n,\emptyset} = h_n$ . What is  $Q_{n,\{1,2,\dots,n-1\}}$ ? (b) Show that  $Q_{n,S}$  is quasisymmetric (as defined in 10.262). (c) Use inclusion-exclusion to express  $M_\alpha$  as a linear combination of  $Q_{n,S}$ 's and vice versa. Use this to find a basis for the space of quasisymmetric polynomials consisting of suitable  $Q$ 's.

**10.264.  $Q$ -Expansion of Schur Polynomials.** For each integer partition  $\lambda$  of  $n$ , prove that  $s_\lambda(x_1, \dots, x_N) = \sum_{U \in \text{SYT}(\lambda)} Q_{n, \text{Des}(U)}(x_1, \dots, x_N)$ , where  $Q_{n,S}$  is defined in 10.263.

## Notes

Macdonald's book [89] contains a comprehensive treatment of symmetric polynomials, with a heavy emphasis on algebraic methods. A more combinatorial development is given by Stanley [127, Chpt. 7]; see the references to that chapter for an extensive bibliography of the literature in this area. Two other relevant references are Fulton [46], which treats tableaux and their connections to representation theory and geometry, and Sagan [121], which explains the role of symmetric polynomials in the representation theory of symmetric groups.

The bijective proof of 10.33 is due to Bender and Knuth [9]. The algorithmic proof of the existence part of the fundamental theorem of symmetric polynomials (outlined in 10.211) is usually attributed to Waring [130]. Some of the seminal papers by Robinson, Schensted, and Knuth on what is now called the RSK correspondence are [79, 116, 122]. The symmetry property 10.112 was first proved by Schützenberger [124], but the combinatorial proof using shadow lines is due to Viennot [135]. Quasisymmetric polynomials (see 10.262) were introduced by Gessel [51].

## Abaci and Antisymmetric Polynomials

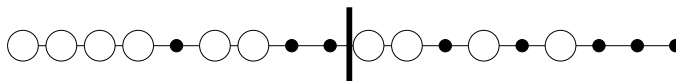
In the last chapter, we used combinatorial operations on tableaux to establish algebraic properties of Schur polynomials and symmetric polynomials. This chapter investigates the interplay between the combinatorics of abaci and the algebraic properties of antisymmetric polynomials. These concepts will be used to establish additional facts about integer partitions and symmetric polynomials. In particular, we will derive some formulas for expanding skew Schur polynomials in terms of various bases.

### 11.1 Abaci and Integer Partitions

An *abacus* is an instrument used in ancient times for performing arithmetical calculations. The abacus consists of one or more runners that contain sliding beads. The following combinatorial object gives a mathematical model of an abacus.

**11.1. Definition: One-Runner Abacus.** An *abacus with one runner* is a function  $w : \mathbb{Z} \rightarrow \{0, 1\}$  such that for some  $m, n$ ,  $w_i = 1$  for all  $i \leq m$  and  $w_i = 0$  for all  $i \geq n$ . We think of  $w$  as an infinite word  $\cdots w_{-2}w_{-1}\underline{w_0}w_1w_2w_3\cdots$  that begins with an infinite string of 1's and ends with an infinite string of 0's. Each 1 is called a *bead*, and each 0 is called a *gap*. Let  $\text{Abc}$  denote the set of all one-runner abaci. An abacus  $w$  is called *justified at position*  $m$  iff  $w_i = 1$  for all  $i \leq m$  and  $w_i = 0$  for all  $i > m$ . Intuitively, an abacus is justified iff all the beads have been pushed to the left as far as they will go. The *weight* of an abacus  $w$ , denoted  $\text{wt}(w)$ , is the number of pairs  $i < j$  with  $w_i < w_j$  (or equivalently,  $w_i = 0$  and  $w_j = 1$ ).

**11.2. Example.** Here is a picture of a one-runner abacus:



This picture corresponds to the mathematical abacus

$$w = \cdots 111101100\underline{1}10101000 \cdots,$$

where the underlined 1 is  $w_0$ . All positions to the left of the displayed region contain beads, and all positions to the right contain gaps.

Consider the actions required to transform  $w$  into a justified abacus. We begin with the bead following the leftmost gap, which slides one position to the left, producing

$$w' = \cdots 111110100\underline{1}10101000 \cdots.$$

The next bead now slides into the position vacated by the previous bead, producing

$$w'' = \cdots 111111000\underline{1}10101000 \cdots.$$

The next bead moves 3 positions to the left to give the abacus

$$w^{(3)} = \cdots 1111111000\underline{1}0101000 \cdots.$$

In the next three steps, the remaining beads move left by 3, 4, and 5 positions, respectively, leading to the abacus

$$w^* = \cdots 111111111\underline{1}00000000 \cdots,$$

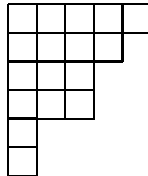
which is justified at position 0. If we list the number of positions that each bead moved, we obtain a weakly increasing sequence:  $1 \leq 1 \leq 3 \leq 3 \leq 4 \leq 5$ . This sequence can be identified with the integer partition  $\lambda = (5, 4, 3, 3, 1, 1)$ . Observe that  $\text{wt}(w) = 17 = |\lambda|$ . This example generalizes as follows.

**11.3. Theorem: Partitions vs. Abaci.** Justification of abaci defines a bijection  $J : \text{Abc} \rightarrow \mathbb{Z} \times \text{Par}$  with inverse  $U : \mathbb{Z} \times \text{Par} \rightarrow \text{Abc}$ . If  $J(w) = (m, \lambda)$ , then  $\text{wt}(w) = |\lambda|$ .

*Proof.* Given an abacus  $w$ , let  $n$  be the least integer with  $w_n = 0$  (the position of the leftmost gap), which exists since  $w$  begins with an infinite string of 1's. Since  $w$  ends with an infinite string of 0's, there are only finitely many  $j > n$  with  $w_j = 1$ ; let these indices be  $j_1 < j_2 < \cdots < j_t$ , where  $n < j_1$ . We justify the abacus by moving the bead at position  $j_1$  left  $\lambda_t = j_1 - n$  places. Then we move the bead at position  $j_2$  left  $\lambda_{t-1} = j_2 - (n + 1)$  places. (We subtract  $n + 1$  since the leftmost gap is now at position  $n + 1$ .) In general, at stage  $k$  we move the bead at position  $j_k$  left  $\lambda_{t+1-k} = j_k - (n + k - 1)$  places. After moving all  $t$  beads, we will have a justified abacus with the leftmost gap located at position  $n + t$ . Since  $n < j_1 < j_2 < \cdots < j_t$ , it follows that  $0 < \lambda_t \leq \lambda_{t-1} \leq \cdots \leq \lambda_1$ . We define  $J(w) = (n + t - 1, \lambda)$  where  $\lambda = (\lambda_1, \dots, \lambda_t)$ . For all  $k$ , moving the bead at position  $j_k$  left  $\lambda_{t+1-k}$  places decreases the weight of the abacus by  $\lambda_{t+1-k}$ . Since a justified abacus has weight zero, it follows that the weight of the original abacus is precisely  $\lambda_t + \cdots + \lambda_1 = |\lambda|$ .

$J$  is a bijection because “unjustification” is a two-sided inverse for  $J$ . More precisely, given  $(m, \mu) \in \mathbb{Z} \times \text{Par}$ , we create an abacus  $U(m, \mu)$  as follows. Start with an abacus justified at position  $m$ . Move the rightmost bead to the right  $\mu_1$  places, then move the next bead to the right  $\mu_2$  places, and so on. This process reverses the action of  $J$ .  $\square$

**11.4. Remark: Computing  $U$ .** The unjustification map  $U$  can also be computed using partition diagrams. We can reconstruct the bead-gap sequence in the abacus  $U(m, \mu)$  by traversing the *frontier* of the diagram of  $\mu$  (traveling northeast) and recording a gap (0) for each horizontal step and a bead (1) for each vertical step. For example, if  $\mu = (5, 4, 3, 3, 1, 1)$ , the diagram of  $\mu$  is



and the bead-gap sequence is 01100110101. To obtain the abacus  $w$ , we prepend an infinite string of 1's, append an infinite string of zeroes, and finally use  $m$  to determine which symbol in the resulting string is considered to be  $w_0$ . One readily checks that this procedure produces the same abacus as the map  $U$  in the previous proof. We can also confirm that the map  $U$  is weight-preserving via the following bijection between the cells of the diagram of  $\mu$  and the pairs  $i < j$  with  $w_i = 0$  and  $w_j = 1$ . Starting at a cell  $c$ , travel south to reach a horizontal edge on the frontier (encoded by some  $w_i = 0$ ). Travel east from  $c$  to reach a vertical edge on the frontier (encoded by some  $w_j = 1$  with  $j > i$ ). For example, the cell in the second row and third column of the diagram above corresponds to the marked gap-bead pair in the associated abacus:

$$\cdots 01100\hat{1}110\hat{1}01 \cdots.$$

## 11.2 Jacobi Triple Product Identity

The *Jacobi triple product identity* is a partition identity that has several applications in combinatorics and number theory. We can give a bijective proof of this identity by using cleverly chosen weights on abaci.

**11.5. Theorem: Jacobi Triple Product Identity.** The following equation holds in the ring  $\mathbb{Q}(u)[[q]]$ :

$$\sum_{m \in \mathbb{Z}} q^{m(m+1)/2} u^m = \prod_{n \geq 1} (1 + uq^n) \prod_{n \geq 0} (1 + u^{-1}q^n) \prod_{n \geq 1} (1 - q^n).$$

*Proof.* Since the formal power series  $\prod_{n \geq 1} (1 - q^n)$  is invertible, it suffices to prove the equivalent identity

$$\sum_{m \in \mathbb{Z}} q^{m(m+1)/2} u^m \prod_{n \geq 1} \frac{1}{1 - q^n} = \prod_{n \geq 1} (1 + uq^n) \prod_{n \geq 0} (1 + u^{-1}q^n). \quad (11.1)$$

Let the weight of an integer  $m$  be  $\text{wt}(m) = q^{m(m+1)/2} u^m$ , and let the weight of a partition  $\mu$  be  $q^{|\mu|}$ . Since  $\prod_{n \geq 1} 1/(1 - q^n) = \sum_{\mu \in \text{Par}} q^{|\mu|}$  by 8.17, the left side of (11.1) is

$$\sum_{(m, \mu) \in \mathbb{Z} \times \text{Par}} \text{wt}(m) \text{wt}(\mu),$$

which is the generating function for the weighted set  $\mathbb{Z} \times \text{Par}$ .

On the other hand, let us define new weights on the set  $\text{Abc}$  as follows. Given an abacus  $w$ , let  $N(w) = \{i \leq 0 : w_i = 0\}$  be the set of nonpositive positions in  $w$  not containing a bead, and let  $P(w) = \{i > 0 : w_i = 1\}$  be the set of positive positions in  $w$  containing a bead. Both  $N(w)$  and  $P(w)$  are finite sets. Define

$$\text{wt}(w) = \prod_{i \in N(w)} (u^{-1}q^{|i|}) \prod_{i \in P(w)} (u^1q^i).$$

We can build an abacus by choosing a bead or a gap in each nonpositive position (choosing “bead” all but finitely many times), and then choosing a bead or a gap in each positive position (choosing “gap” all but finitely many times). The generating function for the choice at position  $i \leq 0$  is  $1 + u^{-1}q^{|i|}$ , while the generating function for the choice at position  $i > 0$  is  $1 + u^1q^i$ . By the product rule for weighted sets (see 8.9), the right side of (11.1) is  $\sum_{w \in \text{Abc}} \text{wt}(w)$ .

To complete the proof, it suffices to argue that the justification bijection  $J : \text{Abc} \rightarrow \mathbb{Z} \times \text{Par}$  is weight-preserving. Suppose  $J(w) = (m, \mu)$  for some abacus  $w$ . The map  $J$  converts  $w$  to an abacus  $w^*$ , justified at position  $m$ , by  $|\mu|$  steps in which some bead moves one position to the left. *Claim 1:* The weight of the justified abacus  $w^*$  is  $\text{wt}(m) = u^m q^{m(m+1)/2}$ . We prove this by considering three cases. When  $m = 0$ ,  $N(w^*) = \emptyset = P(w^*)$ , so  $\text{wt}(w^*) = 1 = \text{wt}(0)$ . When  $m > 0$ ,  $N(w^*) = \emptyset$  and  $P(w^*) = \{1, 2, \dots, m\}$ , so

$$\text{wt}(w^*) = u^m q^{1+2+\dots+m} = u^m q^{m(m+1)/2} = \text{wt}(m).$$

When  $m < 0$ ,  $N(w^*) = \{0, -1, -2, \dots, -(|m| - 1)\}$  and  $P(w^*) = \emptyset$ , so

$$\text{wt}(w^*) = u^{-|m|} q^{0+1+2+\dots+(|m|-1)} = u^m q^{|m|(|m|-1)/2} = u^m q^{m(m+1)/2} = \text{wt}(m).$$

*Claim 2:* If we move one bead one step left in a given abacus  $y$ , the  $u$ -weight stays the same and the  $q$ -weight drops by 1. Let  $i$  be the initial position of the moved bead, and let  $y'$  be the abacus obtained by moving the bead to position  $i - 1$ . If  $i > 1$ , then  $N(y') = N(y)$  and  $P(y') = (P(y) \sim \{i\}) \cup \{i - 1\}$ , so  $\text{wt}(y') = \text{wt}(y)/q$  as desired. If  $i \leq 0$ , then  $P(y') = P(y)$  and  $N(y') = (N(y) \sim \{i - 1\}) \cup \{i\}$  (since the  $N$ -set records positions of gaps), and so  $\text{wt}(y') = \text{wt}(y)q^{|i|}/q^{|i-1|} = \text{wt}(y)/q$ . If  $i = 1$ , then  $P(y') = P(y) \sim \{1\}$  and  $N(y') = N(y) \sim \{0\}$ , so the total  $u$ -weight is preserved and the  $q$ -weight still drops by 1. Finally, combining the two claims gives

$$\text{wt}(w) = \text{wt}(w^*)q^{|\mu|} = \text{wt}(m)\text{wt}(\mu) = \text{wt}(J(w)). \quad \square$$

Variations of the preceding proof can be used to establish other partition identities. As an example, we now sketch a bijective proof of Euler's pentagonal number theorem. Unlike our earlier proof in §8.7, the current proof does not use an involution to cancel oppositely signed objects. We remark that Euler's identity also follows by suitably specializing the Jacobi triple product identity.

**11.6. Euler's Pentagonal Number Theorem.** In  $\mathbb{Q}[[q]]$ , we have

$$\prod_{n=1}^{\infty} (1 - q^n) = \sum_{k \in \mathbb{Z}} (-1)^k q^{\frac{3}{2}k^2 - \frac{1}{2}k}.$$

*Proof.* Note first that

$$\prod_{n=1}^{\infty} (1 - q^n) = \prod_{i \geq 1} (1 - q^{3i}) \prod_{i \geq 1} (1 - q^{3i-1}) \prod_{i \geq 1} (1 - q^{3i-2}).$$

It therefore suffices to prove the identity

$$\prod_{i \geq 1} (1 - q^{3i-1}) \prod_{i \geq 1} (1 - q^{3i-2}) = \sum_{k \in \mathbb{Z}} (-1)^k q^{(3k^2-k)/2} \prod_{i \geq 1} \frac{1}{1 - q^{3i}} = \sum_{(k, \mu) \in \mathbb{Z} \times \text{Par}} (-1)^k q^{3|\mu| + (3k^2-k)/2}. \quad (11.2)$$

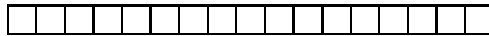
Consider abaci  $w = \{w_{3k+1} : k \in \mathbb{Z}\}$  whose positions are indexed by integers congruent to 1 mod 3. Define  $N(w) = \{i \leq 0 : i \equiv 1 \pmod{3}, w_i = 0\}$  and  $P(w) = \{i > 0 : i \equiv 1 \pmod{3}, w_i = 1\}$ . Let  $\text{sgn}(w) = (-1)^{|N(w)| + |P(w)|}$  and  $\text{wt}(w) = \sum_{i \in N(w) \cup P(w)} |i|$ . We can compute the generating function  $\sum_w \text{sgn}(w)q^{\text{wt}(w)}$  in two ways. On one hand, placing a bead or a gap in each negative position and each positive position leads to the double product on the left side of (11.2). On the other hand, justifying the abacus transforms  $w$  into a pair  $(3k - 2, \mu)$  for some  $k \in \mathbb{Z}$ . As in the proof of the Jacobi triple product identity, one checks that the justified abacus associated to a given integer  $k$  has signed weight  $(-1)^k q^{(3k^2-k)/2}$ , while each of the  $|\mu|$  bead moves in the justification process reduces the  $q$ -weight by 3 and preserves the sign. So the right side of (11.2) is also the generating function for these abaci, completing the proof.  $\square$

### 11.3 Ribbons and $k$ -Cores

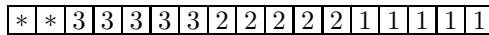
Recall the following fact about division of integers: given integers  $a \geq 0$  and  $k > 0$ , there exist a unique quotient  $q$  and remainder  $r$  satisfying  $a = kq + r$  and  $0 \leq r < k$ . Our next

goal is to develop an analogous operation for dividing an integer partition  $\mu$  by a positive integer  $k$ . The result of this operation will consist of  $k$  “quotient partitions” together with a “remainder partition” with special properties. We begin by describing the calculation of the remainder, which is called a  $k$ -core. Abaci will then be used to establish the uniqueness of the remainder, and this will lead us to the definition of the  $k$  quotient partitions.

To motivate our construction, consider the following pictorial method for performing integer division. Suppose we wish to divide  $a = 17$  by  $k = 5$ , obtaining quotient  $q = 3$  and remainder  $r = 2$ . To find these answers geometrically, first draw a row of 17 boxes:



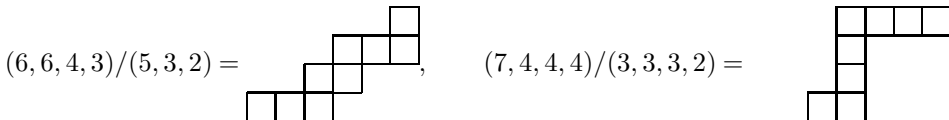
Now, starting at the right end, repeatedly remove strings of five consecutive cells until this is no longer possible. We depict this process by placing an  $i$  in every cell removed at stage  $i$ , and writing a star in any leftover cells:



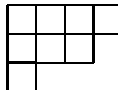
The quotient  $q$  is the number of 5-cell blocks we removed (here 3), and the remainder  $r$  is the number of leftover cells (here 2). This geometric procedure corresponds to the algebraic process of subtracting  $k$  from  $a$  repeatedly until a remainder less than  $k$  is reached. For the purposes of partition division, we now introduce a two-dimensional version of this strip-removal process.

**11.7. Definition: Ribbons.** A *ribbon* is a skew shape that can be formed by starting at a given square, repeatedly moving left or down one step at a time, and including all squares visited in this way. A ribbon consisting of  $k$  cells is called a  $k$ -*ribbon*. A *border ribbon* of a partition  $\mu$  is a ribbon  $R$  contained in  $\text{dg}(\mu)$  such that  $\text{dg}(\mu) \sim R$  is also a partition diagram.

**11.8. Example.** Here are two examples of ribbons:



The first ribbon is a 9-ribbon and a border ribbon of  $(6, 6, 4, 3)$ . The partition  $(4, 3, 1)$  with diagram



has exactly eight border ribbons, four of which begin at the cell  $(1, 4)$ .

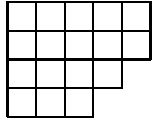
**11.9. Definition:  $k$ -cores.** Let  $k$  be a positive integer. An integer partition  $\nu$  is called a  $k$ -*core* iff no border ribbon of  $\nu$  is a  $k$ -ribbon.

For example,  $(4, 3, 1)$  is a 5-core, but not a  $k$ -core for any  $k < 5$ .

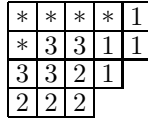
Suppose  $\mu$  is any partition and  $k$  is a positive integer. If  $\mu$  has no border ribbons of size  $k$ , then  $\mu$  is a  $k$ -core. Otherwise, we can pick one such ribbon and remove it from the diagram of  $\mu$  to obtain a smaller partition diagram. We can iterate this process, repeatedly removing a border  $k$ -ribbon from the current partition diagram until this is no longer possible. Since the number of cells decreases at each step, the process will eventually terminate. The final partition  $\nu$  (which may be empty) must be a  $k$ -core. This partition is the “remainder” when  $\mu$  is divided by  $k$ .



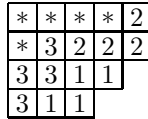
**11.10. Example.** Consider the partition  $\mu = (5, 5, 4, 3)$  with diagram



Let us divide  $\mu$  by  $k = 4$ . We record the removal of border 4-ribbons by entering an  $i$  in each square that is removed at stage  $i$ . Any leftover squares at the end are marked by a star. One possible removal sequence is the following:



Another possible sequence is:



Notice that the three 4-ribbons removed were different, but the final 4-core was the same, namely  $\nu = (4, 1)$ .

We want to show that the  $k$ -core obtained when dividing  $\mu$  by  $k$  depends only on  $\mu$  and  $k$ , not on the choice of which border  $k$ -ribbon is removed at each stage. We now use abaci to prove this result.

**11.11. Definition: Abacus with  $k$  Runners.** A  $k$ -runner abacus is an ordered  $k$ -tuple of abaci. The set of all such objects is denoted  $\text{Abc}^k$ .

**11.12. Theorem: Decimation of Abaci.** For each  $k \geq 1$ , there are mutually inverse bijections  $D_k : \text{Abc} \rightarrow \text{Abc}^k$  (decimation) and  $I_k : \text{Abc}^k \rightarrow \text{Abc}$  (interleaving).

*Proof.* Given  $w = (w_i : i \in \mathbb{Z}) \in \text{Abc}$ , set  $D_k(w) = (w^0, w^1, \dots, w^{k-1})$ , where

$$w^r = (w_{qk+r} : q \in \mathbb{Z}) \quad (0 \leq r < k).$$

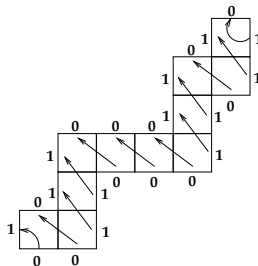
Thus, the abacus  $w^r$  is obtained by reading every  $k$ th symbol in the original abacus (in both directions), starting at position  $r$ . It is routine to check that each  $w^r$  is an abacus. The inverse map interleaves these abaci to reconstruct the original one-runner abacus. More precisely, given  $v = (v^0, v^1, \dots, v^{k-1}) \in \text{Abc}^k$ , let  $I_k(v) = z$  where  $z_{qk+r} = v_q^r$  for all  $q \in \mathbb{Z}$  and  $0 \leq r < k$ . One readily checks that  $I_k(v)$  is an abacus and that  $D_k$  and  $I_k$  are two-sided inverses.  $\square$

By computing  $D_k(U(-1, \mu))$ , we can convert any partition into a  $k$ -runner abacus. We now show that moving one bead left one step on a  $k$ -runner abacus corresponds to removing a border  $k$ -ribbon from the associated partition diagram.

**11.13. Theorem: Bead Motion vs. Ribbon Removal.** Suppose a partition  $\mu$  is encoded by a  $k$ -runner abacus  $w = (w^0, w^1, \dots, w^{k-1})$ . Suppose that  $v$  is a  $k$ -runner abacus obtained from  $w$  by changing one substring  $\dots 01 \dots$  to  $\dots 10 \dots$  in some  $w^i$ . Then the partition  $\nu$  associated to  $v$  can be obtained by removing one border  $k$ -ribbon from  $\mu$ . Moreover, there is a bijection between the set of removable border  $k$ -ribbons in  $\mu$  and the set of occurrences of the substring  $01$  in the components of  $w$ .

*Proof.* Recall from 11.4 that we can encode the frontier of a partition  $\mu$  by writing a 0 (gap) for each horizontal step and writing a 1 (bead) for each vertical step. The word so obtained (when preceded by 1's and followed by 0's) is the 1-runner abacus associated to this partition, and  $w$  is the  $k$ -decimation of this abacus.

Let  $R$  be a border  $k$ -ribbon of  $\mu$ . The southeast border of  $R$ , which is part of the frontier of  $\mu$ , gets encoded as a string of  $k + 1$  symbols  $r_0, r_1, \dots, r_k$ , where  $r_0 = 0$  and  $r_k = 1$ . For instance, the first ribbon in 11.8 has southeast border 0001010011. Note that the *northwest* border of this ribbon is encoded by 1001010010, which is the string obtained by interchanging the initial 0 and the terminal 1 in the original string. The following picture suggests why this property holds for general  $k$ -ribbons.



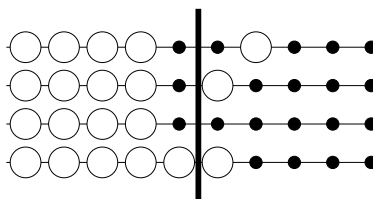
Since  $r_0 = 0$  and  $r_k = 1$  are separated by  $k$  positions in the 1-runner abacus, these two symbols map to two *consecutive* symbols 01 on one of the runners in the  $k$ -runner abacus for  $\mu$ . Changing these symbols to 10 will interchange  $r_0$  and  $r_k$  in the original word. Hence, the portion of the frontier of  $\mu$  consisting of the southeast border of  $R$  gets replaced by the northwest border of  $R$ . So, this bead motion transforms  $\mu$  into the partition  $\nu$  obtained by removing the ribbon  $R$ .

Conversely, each substring 01 in the  $k$ -runner abacus for  $\mu$  corresponds to a unique pair of symbols  $0 \cdots 1$  in the 1-runner abacus that are  $k$  positions apart. This pair corresponds to a unique pair of steps H...V on the frontier that are  $k$  steps apart. Finally, this pair of steps corresponds to a unique removable border  $k$ -ribbon of  $\mu$ . So, the map from these ribbons to occurrences of 01 on the runners of  $w$  is a bijection.  $\square$

**11.14. Example.** Let us convert the partition  $\mu = (5, 5, 4, 3)$  from 11.10 to a 4-runner abacus. First, the 1-runner abacus  $U(-1, \mu)$  is

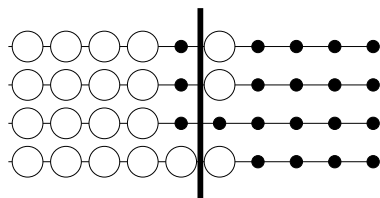
$$\cdots 111000101011000 \cdots$$

Decimating by 4 produces the following 4-runner abacus:

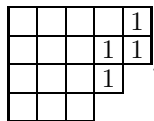


Note that the bead-gap pattern in this abacus can be read directly from the frontier of  $\mu$  by filling in the runners one column at a time, working from left to right. For the purposes of ribbon removal, one may decide arbitrarily where to place the gap corresponding to the first step of the frontier; this decision determines the integer  $m$  in the expression  $U(m, \mu)$ .

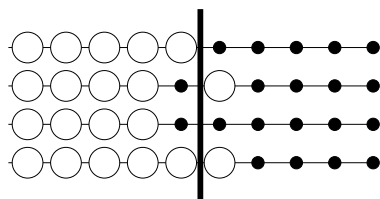
Now let us start removing ribbons. Suppose we push the rightmost bead on the top runner left one position, producing the following abacus:



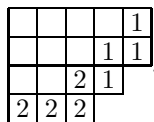
Reading down columns to recover the frontier of the new partition, we obtain the partition  $\nu = (4, 3, 3, 3)$ . We get  $\nu$  from  $\mu$  by removing one border 4-ribbon, as shown.



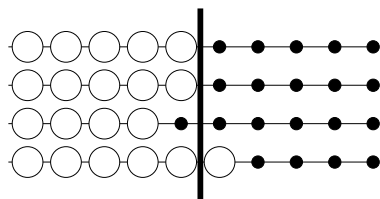
Pushing the same bead one more step on its runner produces the following abacus:



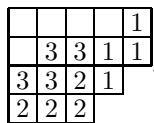
The new partition is  $(4, 3, 2)$ , which arises by removing one border 4-ribbon from  $\nu$ :



Finally, we push the rightmost bead on the second runner left one position to get the following abacus:

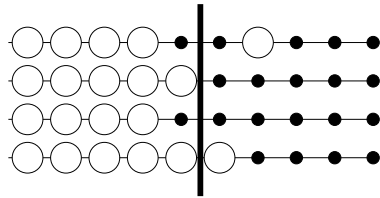


The associated partition is  $(4, 1)$ , as shown here:

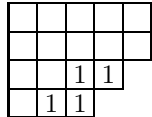


At this point, all runners on the abacus are justified, so no further bead motion is possible. This reflects the fact that we can remove no further border 4-ribbons from the 4-core  $(4, 1)$ .

Now return to the original partition  $\mu$  and the associated 4-runner abacus. Suppose we start by moving the bead on the second runner left one position, producing the following abacus:

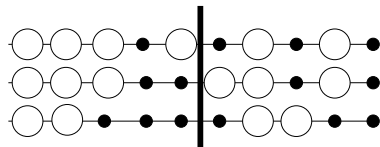


This corresponds to removing a different border 4-ribbon from  $\mu$ :

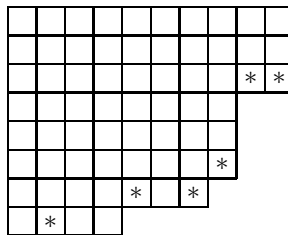


Observe that  $\mu$  has exactly two removable border 4-ribbons, whereas the 4-runner abacus for  $\mu$  has exactly two movable beads, in accordance with the last assertion of 11.13.

**11.15. Example.** Consider the following 3-runner abacus:



We count six beads on this abacus that can be moved one position left without bumping into another bead. Accordingly, we expect the associated partition to have exactly six removable border 3-ribbons. This is indeed the case, as shown below (we have marked the southwestmost cell of each removable ribbon with an asterisk):

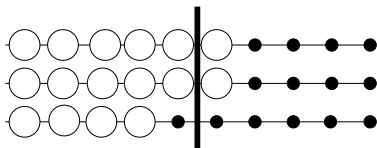


Now we can prove that the  $k$ -core obtained from a partition  $\mu$  by repeated removal of border ribbons is uniquely determined by  $\mu$  and  $k$ .

**11.16. Theorem: Uniqueness of  $k$ -cores.** Suppose  $\mu$  is an integer partition and  $k \geq 1$  is an integer. There is exactly one  $k$ -core  $\rho$  obtainable from  $\mu$  by repeatedly removing border  $k$ -ribbons. We call  $\rho$  the  $k$ -core of  $\mu$ .

*Proof.* Let  $w$  be a fixed  $k$ -runner abacus associated to  $\mu$  (say  $w = D_k(U(-1, \mu))$  for definiteness). As we have seen, a particular sequence of ribbon-removal operations on  $\mu$  corresponds to a particular sequence of bead motions on  $w$ . The operations on  $\mu$  terminate when we reach a  $k$ -core, whereas the corresponding operations on  $w$  terminate when the beads on all runners of  $w$  have been justified. Now  $\rho$  is uniquely determined by the justified  $k$ -runner abacus by applying  $I_k$  and then  $J$ . The key observation is that the justified abacus obtained from  $w$  does not depend on the order in which individual bead moves were made. Thus, the  $k$ -core  $\rho$  does not depend on the order in which border ribbons are removed from  $\mu$ .  $\square$

**11.17. Example.** The theorem shows that we can calculate the  $k$ -core of  $\mu$  by justifying any  $k$ -runner abacus associated to  $\mu$ . For example, consider the partition  $\mu = (10, 10, 10, 8, 8, 8, 7, 4)$  from 11.15. Justifying the 3-runner abacus in that example produces the following abacus:



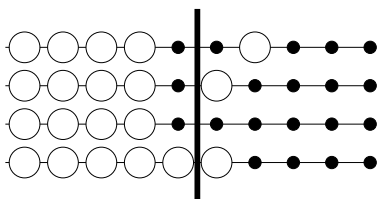
We find that the 3-core of  $\mu$  is  $(1, 1)$ .

## 11.4 $k$ -Quotients and Hooks

Each runner of a  $k$ -runner abacus can be regarded as a one-runner abacus, which corresponds (under the justification bijection  $J$ ) to an element of  $\mathbb{Z} \times \text{Par}$ . This observation leads to the definition of the  $k$ -quotients of a partition.

**11.18. Definition:  $k$ -quotients of a partition.** Let  $\mu$  be a partition and  $k \geq 1$  an integer. Consider the  $k$ -runner abacus  $(w^0, w^1, \dots, w^{k-1}) = D_k(U(-1, \mu))$ . Write  $J(w^i) = (m_i, \nu^i)$  for  $0 \leq i < k$ . The partitions appearing in the  $k$ -tuple  $(\nu^0, \nu^1, \dots, \nu^{k-1})$  are called the  $k$ -quotients of  $\mu$ .

**11.19. Example.** Let  $\mu = (5, 5, 4, 3)$ . In 11.14, we computed the 4-runner abacus  $D_4(U(-1, \mu))$ :



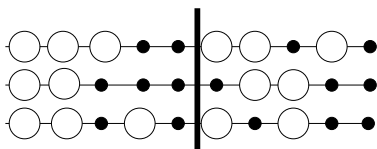
Justifying each runner and converting the resulting 4-runner abacus back to a partition produces the 4-core of  $\mu$ , namely  $(4, 1)$ . On the other hand, converting each runner to a separate partition produces the 4-tuple of 4-quotients of  $\mu$ , namely:

$$((2), (1), (0), (0)).$$

**11.20. Example.** Consider the partition  $\mu = (10, 10, 10, 8, 8, 8, 7, 4)$  from 11.15. We compute

$$U(-1, \mu) = \cdots 111100001000\underline{1}01110011100 \cdots.$$

Decimation by 3 produces the 3-runner abacus shown here:



Justifying each runner shows that the 3-core of  $\mu$  is  $\rho = (1, 1)$ . On the other hand, by regarding each runner separately as a partition, we obtain the 3-tuple of 3-quotients of  $\mu$ :

$$(\nu^0, \nu^1, \nu^2) = ((3, 2, 2), (4, 4), (3, 2, 1)).$$

Observe that  $|\mu| = 65 = 2 + 3 \cdot (7 + 8 + 6) = |\rho| + 3|\nu^0| + 3|\nu^1| + 3|\nu^2|$ .

Now consider what would have happened if we had performed similar computations on the 3-runner abacus for  $\mu$  displayed in 11.15, which is  $D_3(U(0, \mu))$ . The 3-core coming from this abacus is still  $(1, 1)$ , but converting each runner to a partition produces the following 3-tuple:

$$((3, 2, 1), (3, 2, 2), (4, 4)).$$

This 3-tuple arises by cyclically shifting the previous 3-tuple one step to the right. One can check that this holds in general: if the  $k$ -quotients for  $\mu$  are  $(\nu^0, \dots, \nu^{k-1})$ , then the  $k$ -quotients computed using  $D_k(U(m, \mu))$  will be  $(\nu^{k-m'}, \dots, \nu^{k-1}, \nu^0, \nu^1, \dots)$ , where  $m'$  is the integer remainder when  $m + 1$  is divided by  $k$ .

**11.21. Remark.** Here is a way to compute  $(w^0, w^1, \dots, w^{k-1}) = D_k(U(-1, \mu))$  from the frontier of  $\mu$  without writing down the intermediate abacus  $U(-1, \mu)$ . Draw a line of slope  $-1$  starting at the northwest corner of the diagram of  $\mu$ . The first step on the frontier of  $\mu$  lying northeast of this line corresponds to position 0 of the zeroth runner  $w^0$ . The next step is position 0 on  $w^1$ , and so on. The step just southwest of the diagonal line is position  $-1$  on  $w^{k-1}$ , the previous step is position  $-1$  on  $w^{k-2}$ , and so on. To see that this works, one must check that the first step northeast of the diagonal line gets mapped to position 0 on the one-runner abacus  $U(-1, \mu)$ ; we leave this as an exercise.

**11.22. Theorem: Partition Division.** Let  $\text{Core}(k)$  be the set of all  $k$ -cores. There is a bijection

$$\Delta_k : \text{Par} \rightarrow \text{Core}(k) \times \text{Par}^k$$

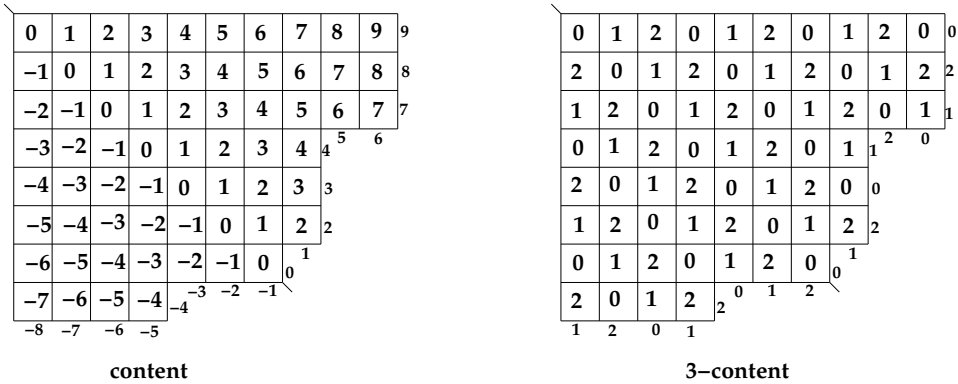
such that  $\Delta_k(\mu) = (\rho, \nu^0, \dots, \nu^{k-1})$ , where  $\rho$  is the  $k$ -core of  $\mu$  and the  $\nu^i$  are the  $k$ -quotients of  $\mu$ . We have

$$|\mu| = |\rho| + k \sum_{i=0}^{k-1} |\nu^i|.$$

*Proof.* The function  $\Delta_k$  is well-defined and maps into the stated codomain. To see that this function is a bijection, we describe its inverse. Given  $(\rho, \nu^0, \dots, \nu^{k-1}) \in \text{Core}(k) \times \text{Par}^k$ , first compute the  $k$ -runner abacus  $(w^0, \dots, w^{k-1}) = D_k(U(-1, \rho))$ . Each  $w^i$  is itself a justified one-runner abacus because  $\rho$  is a  $k$ -core; say  $w^i$  is justified at position  $m_i$ . Now replace each  $w^i$  by  $\nu^i = U(m_i, w^i)$ . Finally, let  $\mu$  be the unique partition satisfying  $J(I_k(\nu^0, \dots, \nu^{k-1})) = (-1, \mu)$ . This construction reverses the one used to produce  $k$ -cores and  $k$ -quotients, so  $\mu$  is the unique partition mapped to  $(\rho, \nu^0, \dots, \nu^{k-1})$  by  $\Delta_k$ .

To prove the formula for  $|\mu|$ , consider the bead movements used to justify the runners of the  $k$ -runner abacus  $D_k(U(-1, \mu))$ . On one hand, every time we move a bead one step left on this abacus, the area of  $\mu$  drops by  $k$  since the bead motion removes one border  $k$ -ribbon. When we finish moving all the beads, we are left with the  $k$ -core  $\rho$ . It follows that  $|\mu| = |\rho| + km$  where  $m$  is the total number of bead motions on all  $k$  runners. On the other hand, for  $0 \leq i < k$ , let  $m_i$  be the number of times we move a bead one step left on runner  $i$ . Then  $m = m_0 + m_1 + \dots + m_{k-1}$ , whereas  $m_i = |\nu^i|$  by 11.3. Substituting these expressions into  $|\mu| = |\rho| + km$  gives the desired formula.  $\square$

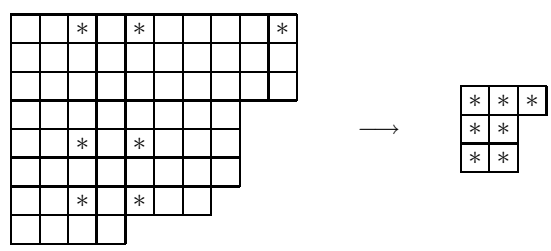
We close our discussion of partition division by describing a way to compute the  $k$ -quotients of  $\mu$  directly from the diagram of  $\mu$ , without recourse to abaci. We will need the following device for labeling cells of  $\text{dg}(\mu)$  and steps on the frontier of  $\mu$  by integers in  $\{0, 1, \dots, k-1\}$ .



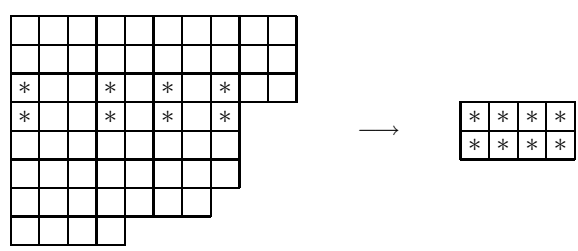
**FIGURE 11.1**  
Content and 3-content of cells and steps.

**11.23. Definition: Content and  $k$ -Content.** Consider a partition diagram for  $\mu$ , drawn with the longest row on top. Introduce a coordinate system so that the northwest corner of the diagram is  $(0, 0)$  and  $(i, j)$  is located  $i$  steps south and  $j$  steps east of the origin. The *content* of the point  $(i, j)$  is  $c(i, j) = j - i$ . The content of a cell in the diagram of  $\mu$  is the content of its southeast corner. The content of a frontier step from  $(i, j)$  to  $(i, j + 1)$  is  $j - i$ . The content of a frontier step from  $(i, j)$  to  $(i - 1, j)$  is  $j - i$ . If  $z$  is a point, cell, or step in the diagram, then the  $k$ -content  $c_k(z)$  is the unique value  $r \in \{0, 1, \dots, k - 1\}$  such that  $c(z) \equiv r \pmod{k}$ .

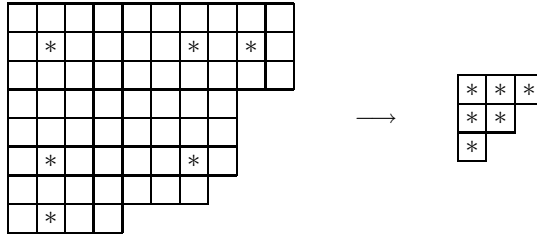
**11.24. Example.** The left side of Figure 11.1 shows the diagram of the partition  $\mu = (10, 10, 10, 8, 8, 8, 7, 4)$  with each cell and frontier step labeled by its content. On the right side of the figure, each cell and step is labeled by its 3-content. Given a cell in the diagram of  $\mu$ , we obtain an associated pair of steps on the frontier of  $\mu$  by traveling south (resp. east) from the cell in question. Suppose we mark all cells whose associated steps both have content zero. Then erase all other cells and shift the marked cells up and left as far as possible. The following diagram results:



This partition  $(3, 2, 2)$  is precisely the zeroth 3-quotient of  $\mu$ . Similarly, marking the cells whose associated steps both have 3-content equal to 1 produces the next 3-quotient of  $\mu$ :



Finally, marking the cells whose associated steps both have 3-content equal to 2 produces the last 3-quotient of  $\mu$ :



In general, to obtain the  $i$ th  $k$ -quotient  $\nu^i$  of  $\mu$  from the diagram of  $\mu$ , label each row (resp. column) of the diagram with the  $k$ -content of the frontier step located in that row (resp. column). Erase all rows and columns not labeled  $i$ . The number of cells remaining in the  $j$ th unerased row is the  $j$ th part of  $\nu^i$ . To see why this works, recall that the cells of  $\nu^i$  correspond bijectively to the pairs of symbols  $0 \cdots 1$  on the  $i$ th runner of the  $k$ -runner abacus for  $\mu$ . In turn, these pairs correspond to pairs of symbols  $w_s = 0, w_t = 1$  on the one-runner abacus for  $\mu$  where  $s < t$  and  $s \equiv i \equiv t \pmod{k}$ . The symbols in positions congruent to  $i \pmod{k}$  come from the steps on the frontier of  $\mu$  whose  $k$ -content is  $i$ . Finally, the relevant pairs of steps on the frontier correspond to the unerased cells in the construction described above. Composing all these bijections, we see that the cells of  $\nu^i$  are in one-to-one correspondence with the unerased cells of the construction. Furthermore, cells in row  $j$  of  $\nu^i$  are mapped onto the unerased cells in the  $j$ th unerased row of  $\mu$ . It follows that the construction at the beginning of this paragraph does indeed produce the  $k$ -quotient  $\nu^i$ .

## 11.5 Antisymmetric Polynomials

We now define antisymmetric polynomials, which form a vector space analogous to the space of symmetric polynomials studied in the last chapter.

**11.25. Definition: Antisymmetric Polynomials.** Let  $K$  be a field containing  $\mathbb{Q}$ . A polynomial  $f \in K[x_1, \dots, x_N]$  is called *antisymmetric* iff for all  $w \in S_N$ ,

$$f(x_{w(1)}, x_{w(2)}, \dots, x_{w(N)}) = \text{sgn}(w)f(x_1, x_2, \dots, x_N).$$

**11.26. Remark.** The group  $S_N$  acts on the set  $\{x_1, \dots, x_N\}$  via  $w \bullet x_i = x_{w(i)}$  for  $w \in S_N$  and  $1 \leq i \leq N$ . This action extends (by the universal mapping property of polynomial rings) to an action of  $S_N$  on  $K[x_1, \dots, x_N]$  such that  $w \bullet f = f(x_{w(1)}, \dots, x_{w(N)})$  for  $w \in S_N$  and  $f \in K[x_1, \dots, x_N]$ . The polynomial  $f$  is antisymmetric iff  $w \bullet f = \text{sgn}(w)f$  for all  $w \in S_N$ . It suffices to check this condition when  $w$  is a basic transposition  $(i, i+1)$ . For, any  $w \in S_N$  can be written as a product of basic transpositions  $w = t_1 t_2 \cdots t_k$ . By hypothesis,  $t_i \bullet (\pm f) = \text{sgn}(t_i)(\pm f) = \mp f$  for all  $i$ , so

$$w \bullet f = t_1 \bullet \cdots \bullet (t_k \bullet f) = (-1)^k f = \text{sgn}(w)f.$$

So  $f \in K[x_1, \dots, x_N]$  is antisymmetric iff

$$f(x_1, \dots, x_{i+1}, x_i, \dots, x_N) = -f(x_1, \dots, x_i, x_{i+1}, \dots, x_N) \quad \text{for all } i < N.$$



**11.27. Example.** The polynomial  $f(x_1, \dots, x_N) = \prod_{1 \leq j < k \leq N} (x_j - x_k)$  is antisymmetric. To check this, consider what happens to the factors in the product when we interchange  $x_i$  and  $x_{i+1}$ . Factors not involving  $x_i$  or  $x_{i+1}$  are unchanged; factors of the form  $(x_i - x_k)$  with  $k > i + 1$  get interchanged with factors of the form  $(x_{i+1} - x_k)$ ; and factors of the form  $(x_j - x_i)$  with  $j < i$  get interchanged with factors of the form  $(x_j - x_{i+1})$ . Finally, the factor  $(x_i - x_{i+1})$  becomes  $(x_{i+1} - x_i) = -(x_i - x_{i+1})$ . Thus,  $(i, i + 1) \bullet f = -f$  for all  $i < N$ , proving antisymmetry of  $f$ .

We remark that the polynomial  $f = \prod_{j < k} (x_j - x_k)$  in the previous example is the *Vandermonde determinant*

$$\det \|x_j^{N-i}\|_{1 \leq i, j \leq N} = \sum_{w \in S_N} \operatorname{sgn}(w) \prod_{i=1}^N x_{w(i)}^{N-i}.$$

(see §12.9 for a combinatorial proof of this assertion). We can use analogous determinants to manufacture additional examples of antisymmetric polynomials.

**11.28. Definition: Monomial Antisymmetric Polynomials.** Let  $\mu = (\mu_1 > \mu_2 > \dots > \mu_N)$  be a strictly decreasing sequence of  $N$  nonnegative integers. Define a polynomial  $a_\mu(x_1, \dots, x_N)$  by the formula

$$a_\mu(x_1, \dots, x_N) = \det \|x_j^{\mu_i}\|_{1 \leq i, j \leq N} = \sum_{w \in S_N} \operatorname{sgn}(w) \prod_{i=1}^N x_{w(i)}^{\mu_i}.$$

We call  $a_\mu$  a *monomial antisymmetric polynomial indexed by  $\mu$* .

To see that  $a_\mu$  really is antisymmetric, note that interchanging  $x_k$  and  $x_{k+1}$  has the effect of interchanging columns  $k$  and  $k + 1$  in the determinant defining  $a_\mu$ . By 9.47, this column switch will change the sign of  $a_\mu$ , as required.

**11.29. Example.** Let  $N = 3$  and  $\mu = (5, 4, 2)$ . Then

$$a_\mu(x_1, x_2, x_3) = +x_1^5 x_2^4 x_3^2 + x_1^4 x_2^5 x_3^2 + x_1^2 x_2^5 x_3^4 - x_1^4 x_2^5 x_3^2 - x_1^5 x_2^2 x_3^4 - x_1^2 x_2^4 x_3^5.$$

As the previous example shows,  $a_\mu(x_1, \dots, x_N)$  is a sum of  $N!$  distinct monomials obtained by rearranging the subscripts (or equivalently, the exponents) in the monomial  $x_1^{\mu_1} x_2^{\mu_2} \dots x_N^{\mu_N}$ . Each monomial appears in the sum with sign  $+1$  or  $-1$ , where the sign of  $x_1^{e_1} \dots x_N^{e_N}$  depends on the parity of the number of basic transpositions needed to transform the sequence  $(e_1, \dots, e_N)$  to the sorted sequence  $(\mu_1, \dots, \mu_N)$ . It follows from these remarks that  $a_\mu$  is a nonzero homogeneous polynomial of degree  $|\mu| = \mu_1 + \dots + \mu_N$ .

**11.30. Definition:  $\delta(N)$ .** For each  $N \geq 1$ , let  $\delta(N) = (N - 1, N - 2, \dots, 2, 1, 0)$ .

The *strictly* decreasing sequences  $\mu = (\mu_1 > \mu_2 > \dots > \mu_N)$  correspond bijectively to the *weakly* decreasing sequences  $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N)$  via the maps  $\mu \mapsto \mu - \delta(N)$  and  $\lambda \mapsto \lambda + \delta(N)$ . It follows that each polynomial  $a_\mu$  can be written  $a_{\lambda + \delta(N)}$  for a unique partition  $\lambda \in \operatorname{Par}_N$ . This indexing scheme will be used frequently below. Note that when  $\lambda = (0, \dots, 0)$ , we have  $\mu = \delta(N)$  and  $a_{\delta(N)} = \prod_{1 \leq j < k \leq N} (x_j - x_k)$  (see 11.27). Observe that  $a_{\delta(N)}$  is a homogeneous polynomial of degree  $N(N - 1)/2 = \binom{N}{2}$ .

**11.31. Definition: Spaces of Antisymmetric Polynomials.** For a given field  $K$ , let  $A_N$  be the set of all antisymmetric polynomials in  $K[x_1, \dots, x_N]$ . Let  $A_N^n$  consist of those polynomials in  $A_N$  that are homogeneous of degree  $n$ , together with the zero polynomial.

One readily verifies that  $A_N$  is a vector subspace of  $K[x_1, \dots, x_N]$ , and each  $A_N^n$  is a subspace of  $A_N$ . We now exhibit bases for these vector spaces involving monomial antisymmetric polynomials. We use the notation  $\text{Par}_N^d(n)$  to denote the set of all partitions of  $n$  into  $N$  distinct nonnegative parts, and  $\text{Par}_N^d = \bigcup_{n \geq \binom{N}{2}} \text{Par}_N^d(n)$ .

**11.32. Theorem: Monomial Basis for  $A_N^n$ .** Assume  $K$  is a field containing  $\mathbb{Q}$ . If  $n < \binom{N}{2}$ , then  $A_N^n = \{0\}$ . If  $n \geq \binom{N}{2}$ , then

$$\{a_\mu : \mu \in \text{Par}_N^d(n)\} = \{a_{\lambda+\delta(N)} : \lambda \in \text{Par}_N(n - N(N-1)/2)\}$$

is a basis of the  $K$ -vector space  $A_N^n$ . Hence, the collection

$$\{a_\mu : \mu \in \text{Par}_N^d\} = \{a_{\lambda+\delta(N)} : \lambda \in \text{Par}_N\}$$

is a basis of  $A_N$ .

*Proof.* Suppose  $e = (e_1, \dots, e_N)$  is any exponent sequence,  $f \in A_N$  is an arbitrary antisymmetric polynomial, and  $w \in S_N$ . Let  $c$  be the coefficient of  $x^e = x_1^{e_1} \cdots x_N^{e_N}$  in  $f$ , so

$$f = cx_1^{e_1} \cdots x_N^{e_N} + \text{other terms.}$$

Acting by  $w$ , we see that

$$\begin{aligned} \text{sgn}(w)f = w \bullet f &= cx_{w(1)}^{e_1} \cdots x_{w(N)}^{e_N} + \text{other terms} \\ &= cx_1^{e_{w^{-1}(1)}} \cdots x_N^{e_{w^{-1}(N)}} + \text{other terms} \\ &= cx^{w*e} + \text{other terms,} \end{aligned}$$

where  $w * e = (e_{w^{-1}(1)}, \dots, e_{w^{-1}(N)})$ . In other words, writing  $f|_{x^\alpha}$  for the coefficient of  $x^\alpha$  in  $f$ , we have  $f|_{x^{w*e}} = \text{sgn}(w)(f|_{x^e})$ .

Let us apply this fact to an exponent sequence  $e$  such that  $e_i = e_j$  for some  $i \neq j$ . Let  $w = (i, j)$ , so that  $w * e = e$  and  $\text{sgn}(w) = -1$ . It follows that  $c = -c$ , so  $2c = 0$ . Because  $K$  contains  $\mathbb{Q}$ , we deduce  $c = 0$  in  $K$ . This means that no antisymmetric polynomial contains any monomial with a repeated value in its exponent vector. In particular, the smallest possible degree of a monomial that can appear with nonzero coefficient in any antisymmetric polynomial is  $0 + 1 + 2 + \cdots + (N-1) = \binom{N}{2}$ . This proves the first assertion of the theorem.

For the second assertion, recall that  $\lambda \mapsto \lambda + \delta(N)$  is a bijection from  $\text{Par}_N(n - \binom{N}{2})$  to  $\text{Par}_N^d(n)$ . So we need only show that  $\{a_\mu : \mu \in \text{Par}_N^d(n)\}$  is a basis for  $A_N^n$ . To show that this set spans  $A_N^n$ , fix  $f \in A_N^n$ . By the previous paragraph, we can write  $f = \sum_\alpha c_\alpha x^\alpha$  where we sum over all sequences  $(\alpha_1, \dots, \alpha_N) \in \mathbb{N}^N$  with *distinct* entries summing to  $n$ , and each  $c_\alpha$  lies in  $K$ . We claim

$$f = \sum_{\nu \in \text{Par}_N^d(n)} c_\nu a_\nu.$$

To prove this, we check the coefficient of  $x^\alpha$  on each side. Choose  $\mu \in \text{Par}_N^d(n)$  and  $w \in S_N$  such that  $w * \mu = \alpha$  ( $\mu$  consists of the entries of  $\alpha$  sorted into decreasing order). By the first paragraph of the proof,

$$f|_{x^\alpha} = f|_{x^{w*\mu}} = \text{sgn}(w)(f|_{x^\mu}) = \text{sgn}(w)c_\mu.$$

On the other side,  $a_\nu|_{x^\alpha} = 0$  for all  $\nu \neq \mu$  (since no rearrangement of  $\nu$  equals  $\alpha$ ). For  $\nu = \mu$ ,

antisymmetry gives  $a_\mu|_{x^\alpha} = \text{sgn}(w)(a_\mu|_{x^\mu}) = \text{sgn}(w)$ . Multiplying by  $c_\nu$  and summing over all  $\nu$ , the coefficient of  $x^\alpha$  in  $\sum_\nu c_\nu a_\nu$  is  $\text{sgn}(w)c_\mu$ , as desired.

To prove linear independence, suppose

$$0 = \sum_{\nu \in \text{Par}_N^d(n)} d_\nu a_\nu \quad (d_\nu \in K).$$

For a fixed  $\mu \in \text{Par}_N^d(n)$ ,  $a_\mu$  is the only polynomial among the  $a_\nu$ 's that involves the monomial  $x^\mu$ . Extracting this coefficient on both sides of the given equation, we find that  $0 = d_\mu \cdot 1 = d_\mu$ . Since  $\mu$  was arbitrary, all  $d_\mu$ 's are zero.  $\square$

The next result explains the relationship between the various vector spaces  $\Lambda_N^k$  and  $A_N^n$ .

**11.33. Theorem: Symmetric vs. Antisymmetric Polynomials.** For each  $k \geq 0$ , the vector spaces  $\Lambda_N^k$  and  $A_N^{k+\binom{N}{2}}$  are isomorphic, as are the vector spaces  $\Lambda_N$  and  $A_N$ . In each of these cases, an isomorphism is given by the formula  $M(f) = f \cdot a_{\delta(N)}$  for  $f \in \Lambda_N$ , and the inverse isomorphism sends  $g \in A_N$  to  $g/a_{\delta(N)}$ . In particular, every antisymmetric polynomial in  $N$  variables is divisible by the polynomial  $a_{\delta(N)}$ .

*Proof.* Fix  $k \geq 0$ , and consider the map  $M = M_k : \Lambda_N^k \rightarrow K[x_1, \dots, x_N]$  defined by  $M(f) = f \cdot a_{\delta(N)}$  for  $f \in \Lambda_N^k$ . Note, first, that  $f$  is homogeneous of degree  $k$  and  $a_{\delta(N)}$  is homogeneous of degree  $\binom{N}{2}$ , so  $M(f)$  is homogeneous of degree  $k + \binom{N}{2}$ . Second,  $M(f)$  is antisymmetric, since for any  $w \in S_N$ ,

$$w \bullet (f a_{\delta(N)}) = (w \bullet f) \cdot (w \bullet a_{\delta(N)}) = f \cdot (\text{sgn}(w) a_{\delta(N)}) = \text{sgn}(w) (f a_{\delta(N)}).$$

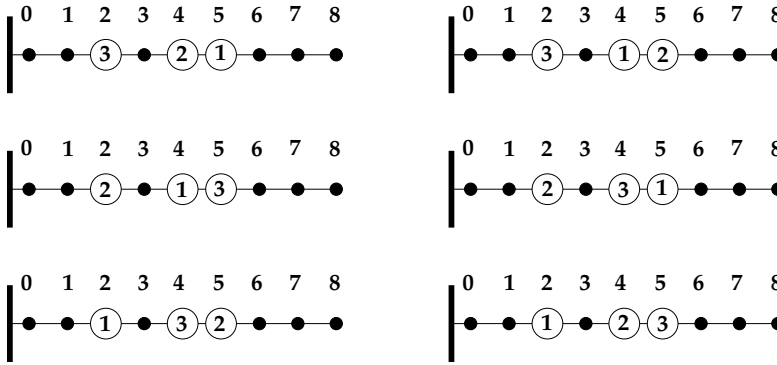
So the map  $M$  takes values in the space  $A_N^{k+\binom{N}{2}}$ . Third, one immediately verifies that  $M$  is a  $K$ -linear map. Fourth, the kernel of this linear map is zero:  $M(f) = 0$  implies  $f \cdot a_{\delta(N)} = 0$ , which implies  $f = 0$  since  $a_{\delta(N)}$  is a nonzero element of the integral domain  $K[x_1, \dots, x_N]$ . So  $M$  is injective. Fifth,  $M$  must also be surjective since its domain and codomain are vector spaces having the same finite dimension  $|\text{Par}_N(k)|$ . So each  $M_k$  is an isomorphism. Since  $\Lambda_N$  (resp.  $A_N$ ) is the direct sum of subspaces  $\Lambda_N^k$  (resp.  $A_N^{k+\binom{N}{2}}$ ), it follows that  $\Lambda_N$  and  $A_N$  are isomorphic as well. Finally, surjectivity of the map  $f \mapsto f a_{\delta(N)}$  means that every antisymmetric polynomial  $g$  has the form  $f a_{\delta(N)}$  for some symmetric polynomial  $f$ . So  $g$  is divisible by  $a_{\delta(N)}$  in  $K[x_1, \dots, x_N]$ .  $\square$

**11.34. Remark.** Suppose we apply the inverse of the isomorphism  $M_k$  to the basis  $\{a_{\lambda+\delta(N)} : \lambda \in \text{Par}_N(k)\}$  of  $A_N^{k+\binom{N}{2}}$ . We will obtain a basis  $\{a_{\lambda+\delta(N)}/a_{\delta(N)} : \lambda \in \text{Par}_N(k)\}$  of  $\Lambda_N^k$ . It turns out that  $a_{\lambda+\delta(N)}/a_{\delta(N)}$  is none other than the Schur polynomial  $s_\lambda(x_1, \dots, x_N)$ ! To prove this fact and other properties of antisymmetric polynomials, we will use the *labeled abaci* introduced below.

## 11.6 Labeled Abaci

Given  $\mu = (\mu_1 > \mu_2 > \dots > \mu_N)$ , recall that the monomial antisymmetric polynomial indexed by  $\mu$  is defined by

$$a_\mu(x_1, \dots, x_N) = \sum_{w \in S_N} \text{sgn}(w) \prod_{i=1}^N x_{w(i)}^{\mu_i}.$$



**FIGURE 11.2**  
Labeled abaci.

The next definition introduces a set of signed, weighted combinatorial objects to model this formula.

**11.35. Definition: Labeled Abaci.** A *labeled abacus with  $N$  beads* is a word  $v = (v_i : i \geq 0)$  such that each letter  $1, \dots, N$  appears exactly once in  $v$ , and all other letters of  $v$  are zero. We think of the indices  $i$  as positions on an abacus containing one runner that extends to infinity in the positive direction. When  $v_i = 0$ , there is a gap at position  $i$  on the abacus; when  $v_i = j > 0$ , there is a bead labeled  $j$  at position  $i$ . The *weight* of the abacus  $v$  is

$$\text{wt}(v) = \prod_{i: v_i > 0} x_{v_i}^i.$$

So if bead  $j$  is located at position  $i$ , this bead contributes a factor of  $x_j^i$  to the weight.

We can encode a labeled abacus by specifying the positions occupied by the beads and the ordering of the bead labels. Formally, define  $\text{pos}(v) = (\mu_1 > \mu_2 > \dots > \mu_N)$  to be the indices  $i$  such that  $v_i > 0$ . Then define  $w(v) = (v_{\mu_1}, \dots, v_{\mu_N}) \in S_N$ . We define the *sign* of  $v$  to be the sign of the permutation  $w(v)$ , which is  $(-1)^{\text{inv}(w(v))}$ . Let  $\text{LAbc}$  be the set of all labeled abaci, and for each  $\mu \in \text{Par}_N^d$ , let

$$\text{LAbc}(\mu) = \{v \in \text{LAbc} : \text{pos}(v) = \mu\}.$$

For each  $\mu \in \text{Par}_N^d$ , there is a bijection between  $\text{LAbc}(\mu)$  and  $S_N$  given by  $v \mapsto w(v)$ . Furthermore, an abacus  $v \in \text{LAbc}(\mu)$  has sign  $\text{sgn}(w(v))$  and weight  $\prod_{i=1}^N x_{w(v)_i}^{\mu_i}$ . So

$$\sum_{v \in \text{LAbc}(\mu)} \text{sgn}(v) \text{wt}(v) = \sum_{w \in S_N} \text{sgn}(w) \prod_{i=1}^N x_{w(i)}^{\mu_i} = a_\mu(x_1, \dots, x_N).$$

**11.36. Example.** Let  $N = 3$  and  $\nu = (5, 4, 2)$ . Earlier, we computed

$$a_\nu(x_1, x_2, x_3) = +x_1^5 x_2^4 x_3^2 + x_1^4 x_2^2 x_3^5 + x_1^2 x_2^5 x_3^4 - x_1^4 x_2^5 x_3^2 - x_1^5 x_2^2 x_3^4 - x_1^2 x_2^4 x_3^5.$$

The six terms in this polynomial come from the six labeled abaci in  $\text{LAbc}(\nu)$  shown in Figure 11.2. Observe that we read labels from right to left in  $v$  to obtain the permutation  $w(v)$ . This is necessary so that the “leading term”  $x_1^{\nu_1} \cdots x_N^{\nu_N}$  will correspond to the identity permutation and have a positive sign.

Informally, we *justify* a labeled abacus  $v \in \text{LAbc}(\mu)$  by moving all beads to the left as far as they will go. This produces a justified labeled abacus  $J(v) = (w_N, \dots, w_2, w_1, 0, 0, \dots) \in \text{LAbc}(\delta(N))$ , where  $(w_N, \dots, w_1) = w(v)$ . To recover  $v$  from  $J(v)$ , first write  $\mu = \lambda + \delta(N)$  for some  $\lambda \in \text{Par}_N$ . Move the rightmost bead (labeled  $w_1$ ) to the right  $\lambda_1$  positions from position  $N - 1$  to position  $N - 1 + \lambda_1 = \mu_1$ . Then move the next bead (labeled  $w_2$ ) to the right  $\lambda_2$  positions from position  $N - 2$  to position  $N - 2 + \lambda_2 = \mu_2$ , and so on.

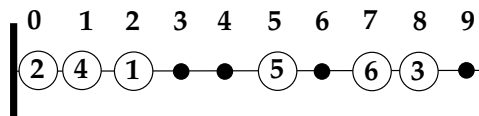
## 11.7 Pieri Rule for $p_k$

The product of an antisymmetric polynomial and a symmetric polynomial is an antisymmetric polynomial (see 11.96), which can be written as a linear combination of the monomial antisymmetric polynomials. In the next few sections, we will derive several “Pieri rules” for expressing a product  $a_{\lambda+\delta(N)}g$  (where  $g$  is symmetric) in terms of the  $a_\mu$ ’s. We begin by considering the case where  $g = p_k(x_1, \dots, x_N) = \sum_{i=1}^N x_i^k$  is a power-sum symmetric polynomial.

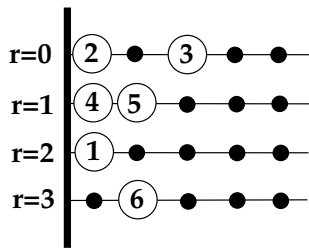
We know  $a_{\lambda+\delta(N)}(x_1, \dots, x_N)$  is a sum of signed terms, each of which represents a labeled abacus with beads in positions given by  $\mu = \lambda + \delta(N)$ . If we multiply some term in this sum by  $x_i^k$ , what happens to the associated abacus? Recalling that the power of  $x_i$  tells us where bead  $i$  is located, we see that this multiplication should move bead  $i$  to the right  $k$  positions. This bead motion occurs all at once, not one step at a time, so bead  $i$  is allowed to “jump over” any beads between its original position and its destination. However, there is a problem if the new position for bead  $i$  already contains a bead. In the proofs below, we will see that two objects of opposite sign cancel whenever a *bead collision* like this occurs. If there is no collision, the motion of bead  $i$  will produce a new labeled abacus whose  $x_i$ -weight has increased by  $k$ . However, the sign of the new abacus (compared to the original) depends on the parity of the number of beads that bead  $i$  “jumps over” when it moves to its new position.

To visualize these ideas more conveniently, we *decimate* our labeled abacus to obtain a labeled abacus with  $k$  runners. Formally, the  $k$ -decimation of the labeled abacus  $v = (v_j : j \geq 0) \in \text{LAbc}(\lambda + \delta(N))$  is the  $k$ -tuple  $(v^0, v^1, \dots, v^{k-1})$ , where  $v_q^r = v_{qk+r}$ . Moving a bead from position  $j$  to position  $j + k$  on the original abacus corresponds to moving a bead one position along its runner on the  $k$ -runner abacus. If there is already a bead in position  $j + k$ , we say that this bead move causes a *bead collision*. Otherwise, the bead motion produces a new labeled abacus in  $\text{LAbc}(\nu + \delta(N))$ , for some  $\nu \in \text{Par}_N$ . By ignoring the labels in the decimated abacus, we see that  $\nu$  arises from  $\lambda$  by adding one  $k$ -ribbon at the border. The shape of this ribbon determines the sign change caused by the bead move, as illustrated in the following example.

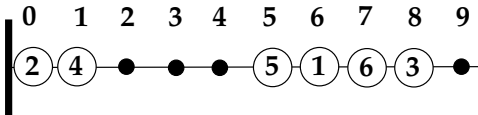
**11.37. Example.** Take  $N = 6$ ,  $k = 4$ ,  $\lambda = (3, 3, 2, 0, 0, 0)$ , and  $\mu = \lambda + \delta(6) = (8, 7, 5, 2, 1, 0)$ . Consider the following labeled abacus  $v$  in  $\text{LAbc}(\mu)$ :



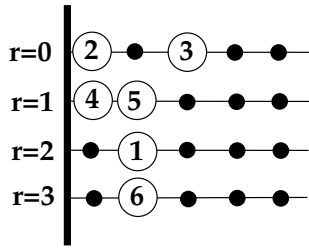
This abacus has weight  $x_1^2 x_2^0 x_3^8 x_4^1 x_5^5 x_6^7$  and sign  $\text{sgn}(3, 6, 5, 1, 4, 2) = (-1)^{10} = +1$ . Decimation by 4 produces the following 4-runner abacus:



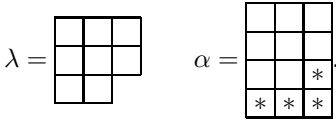
Suppose we move bead 1 four positions to the right in the original abacus, from position 2 to position 6:



The new abacus has weight  $x_1^6 x_2^0 x_3^8 x_4^1 x_5^5 x_6^7 = \text{wt}(v) x_1^4$  and sign  $\text{sgn}(3, 6, 1, 5, 4, 2) = (-1)^9 = -1$ . The change in weight arose since bead 1 moved 4 positions to the right. The change in sign arose since bead 1 passed one other bead (bead 5) to reach its new position, and one basic transposition is needed to transform the permutation 3, 6, 5, 1, 4, 2 into 3, 6, 1, 5, 4, 2. The decimation of the new abacus looks like:

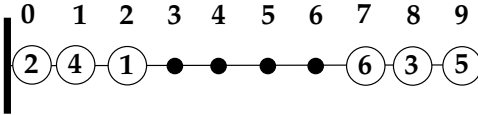


This abacus is in  $\text{LAbc}(\nu) = \text{LAbc}(\alpha + \delta(6))$ , where  $\nu = (8, 7, 6, 5, 1, 0)$  and  $\alpha = (3, 3, 3, 3, 0, 0)$ . Compare the diagrams of the partitions  $\lambda$  and  $\alpha$ :

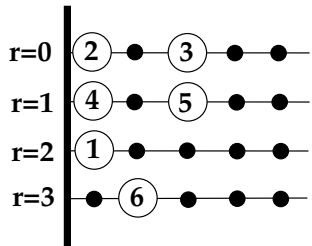


We obtain  $\alpha$  from  $\lambda$  by adding a new border 4-ribbon. To go from  $\lambda$  to  $\alpha$ , we change part of the frontier of  $\lambda$  from NEENE (where the first N step corresponds to bead 1) to EEENN (where the last N step corresponds to bead 1). There is one other N in this string, corresponding to the one bead (labeled 5) that bead 1 passes when it moves to position 6. Thus the number of passed beads (which is 1, here) is one less than the number of rows occupied by the new border ribbon (which is 2, here).

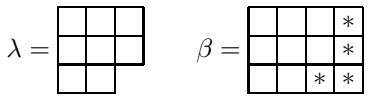
Let us return to the original abacus  $v$  and move bead 5 four positions, from position 5 to position 9:



This abacus has weight  $x_1^2 x_2^0 x_3^8 x_4^1 x_5^9 x_6^7 = \text{wt}(v) x_5^4$  and sign  $\text{sgn}(5, 3, 6, 1, 4, 2) = (-1)^{10} = +1$ . Note that the sign is unchanged since two basic transpositions are required to change the permutation  $3, 6, 5, 1, 4, 2$  into  $5, 3, 6, 1, 4, 2$ . The decimation of the new abacus is:

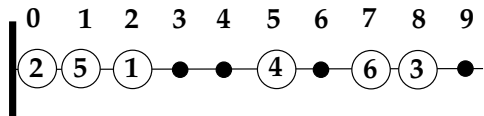


This abacus lies in  $\text{LAbc}(\beta + \delta(6))$  where  $\beta = (4, 4, 4, 0, 0, 0)$ . The diagram of  $\beta$  arises by adding a border 4-ribbon to the diagram of  $\lambda$ :



This time the frontier changed from ...NENNE... (where the first N is bead 5) to ...EENNN... (where the last N is bead 5). The moved bead passed two other beads (beads 3 and 6), which is one less than the number of rows in the new ribbon (three). In general, the number of passed beads is one less than the number of N's in the frontier substring associated to the added ribbon, which is one less than the number of rows in the added ribbon.

Finally, consider what would happen if we tried to move bead 4 (in the original abacus) four positions to the right. A bead collision occurs with bead 5, so this move is impossible. Now consider the labeled abacus  $v'$  obtained by interchanging the labels 4 and 5 in  $v$ :



Moving bead 5 four positions to the right in  $v'$  causes a bead collision with bead 4. Notice that  $\text{sgn}(v') = -\text{sgn}(v)$  since  $[3, 6, 4, 1, 5, 2] = (4, 5) \circ [3, 6, 5, 1, 4, 2]$ . Also note that  $\text{wt}(v) x_4^4 = \text{wt}(v') x_5^4$ ; this equality is valid precisely because of the bead collisions. The abaci  $v$  and  $v'$  are examples of a matched pair of oppositely signed objects that will cancel in the proof of the Pieri rule, given below.

The observations in the last example motivate the following definition.

**11.38. Definition: Spin and Sign of Ribbons.** The *spin* of a ribbon  $R$ , denoted  $\text{spin}(R)$ , is one less than the number of rows occupied by the ribbon. The *sign* of  $R$  is  $\text{sgn}(R) = (-1)^{\text{spin}(R)}$ .

We now have all the combinatorial ingredients needed to prove the Pieri rule for multiplication by a power-sum polynomial.

**11.39. Theorem: Antisymmetric Pieri Rule for  $p_k$ .** For all  $\lambda \in \text{Par}_N$  and all  $k \geq 1$ , the following identity holds in  $K[x_1, \dots, x_N]$ :

$$a_{\lambda + \delta(N)}(x_1, \dots, x_N) p_k(x_1, \dots, x_N) = \sum_{\substack{\beta \in \text{Par}_N: \\ \beta/\lambda \text{ is a } k\text{-ribbon } R}} \text{sgn}(R) a_{\beta + \delta(N)}(x_1, \dots, x_N).$$

*Proof.* Let  $X$  be the set of pairs  $(v, i)$ , where  $v \in \text{LAbc}(\lambda + \delta(N))$  and  $1 \leq i \leq N$ . For  $(v, i) \in X$ , set  $\text{sgn}(v, i) = \text{sgn}(v)$  and  $\text{wt}(v, i) = \text{wt}(v)x_i^k$ . Then  $a_{\lambda + \delta(N)}p_k = \sum_{z \in X} \text{sgn}(z) \text{wt}(z)$ . We introduce a weight-preserving, sign-reversing involution  $I$  on  $X$ . Given  $(v, i)$  in  $X$ , try to move bead  $i$  to the right  $k$  positions in  $v$ . If this move causes a bead collision with bead  $j$ , let  $v'$  be  $v$  with beads  $i$  and  $j$  switched, and set  $I(v, i) = (v', j)$ . Otherwise, set  $I(v, i) = (v, i)$ . One verifies that  $I$  is an involution.

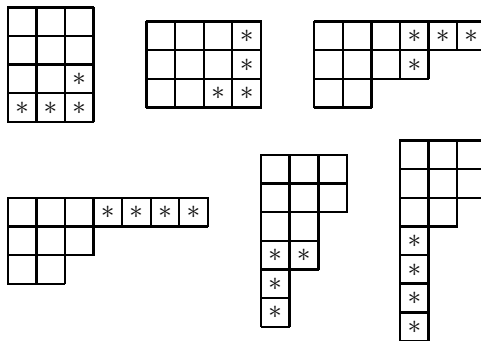
Consider the case where  $I(v, i) = (v', j) \neq (v, i)$ . Since the label permutation  $w(v')$  is obtained from  $w(v)$  by multiplying by the basic transposition  $(i, j)$ ,  $\text{sgn}(v', j) = \text{sgn}(v') = -\text{sgn}(v) = -\text{sgn}(v, i)$ . The weight of  $v$  must have the form  $x_i^a x_j^{a+k} \dots$  because of the bead collision, so  $\text{wt}(v') = x_j^a x_i^{a+k} \dots$ . It follows that  $\text{wt}(v, i) = \text{wt}(v)x_i^k = \text{wt}(v')x_j^k = \text{wt}(v', j)$ . Thus,  $I$  is a weight-preserving, sign-reversing map.

Now consider a fixed point  $(v, i)$  of  $I$ . Let  $v^*$  be the abacus obtained from  $v$  by moving bead  $i$  to the right  $k$  positions, so  $\text{wt}(v^*) = \text{wt}(v)x_i^k = \text{wt}(v, i)$ . Since the unlabeled  $k$ -runner abacus for  $v^*$  arises from the unlabeled  $k$ -runner abacus for  $v$  by moving one bead one step along its runner, it follows that  $v^* \in \text{LAbc}(\beta + \delta(N))$  for a unique  $\beta \in \text{Par}_N$  such that  $R = \beta/\lambda$  is a  $k$ -ribbon. As argued earlier,  $\text{sgn}(v^*)$  differs from  $\text{sgn}(v)$  by  $\text{sgn}(R) = (-1)^{\text{spin}(R)}$ , which is the number of beads that bead  $i$  passes over when it moves. Conversely, any abacus  $y$  counted by  $a_{\beta + \delta(N)}$  (for some shape  $\beta$  as above) arises from a unique fixed point  $(v, i) \in X$ , since the moved bead  $i$  is uniquely determined by the shapes  $\lambda$  and  $\beta$ , and  $v$  is determined from  $y$  by moving the bead  $i$  back  $k$  positions. These remarks show that the sum appearing on the right side of the theorem is the generating function for the fixed point set of  $I$ , which completes the proof.  $\square$

**11.40. Example.** When  $N = 6$ , we calculate

$$a_{(3,3,2)+\delta(6)}p_4 = -a_{(3,3,3,3)+\delta(6)} + a_{(4,4,4)+\delta(6)} - a_{(6,4,2)+\delta(6)} + a_{(7,3,2)+\delta(6)} + a_{(3,3,2,2,1,1)+\delta(6)}$$

by adding border 4-ribbons to the shape  $(3, 3, 2)$ , as shown here:



Observe that the last shape pictured does *not* contribute to the sum because it has more than  $N$  parts. An antisymmetric polynomial indexed by this shape would appear for  $N \geq 7$ .

## 11.8 Pieri Rule for $e_k$

Next we derive Pieri rules for calculating  $a_{\lambda + \delta(N)}e_k$  and  $a_{\lambda + \delta(N)}h_k$ . Our starting point is the following expression for the elementary symmetric polynomial  $e_k$ :

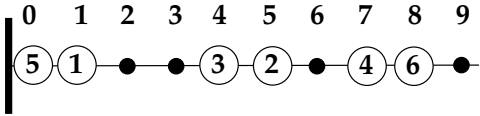
$$e_k(x_1, \dots, x_N) = \sum_{\substack{S \subseteq \{1, 2, \dots, N\} \\ |S| = k}} \prod_{j \in S} x_j.$$



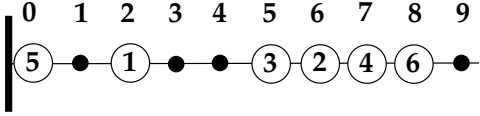
Let  $S = \{j_1, \dots, j_k\}$  be a fixed  $k$ -element subset of  $\{1, 2, \dots, N\}$ . Then  $\prod_{j \in S} x_j = x_{j_1} x_{j_2} \cdots x_{j_k}$  is a typical term in the polynomial  $e_k$ . On the other hand, a typical term in  $a_{\lambda + \delta(N)}$  corresponds to a signed, weighted abacus  $v$ . Let us investigate what happens to the abacus when we multiply such a term by  $x_{j_1} \cdots x_{j_k}$ .

Since the power of  $x_j$  indicates which position bead  $j$  occupies, multiplication by  $x_{j_1} \cdots x_{j_k}$  should cause each of the beads labeled  $j_1, \dots, j_k$  to move one position to the right. We execute this action by scanning the positions of  $v$  from right to left. Whenever we see a bead labeled  $j$  for some  $j \in S$ , we move this bead one step to the right, thus multiplying the weight by  $x_j$ . Bead collisions may occur, which will lead to object cancellations in the proof below. In the case where no bead collisions happen, we obtain a new abacus  $v^* \in a_{\nu + \delta(N)}$ . The beads on this abacus occur in the same order as on  $v$ , so  $w(v^*) = w(v)$  and  $\text{sgn}(v^*) = \text{sgn}(v)$ . Recalling that the parts of  $\lambda$  (resp.  $\nu$ ) count the number of bead moves needed to justify the beads in  $v$  (resp.  $v^*$ ), it follows that  $\nu \in \text{Par}_N$  is a partition obtained from  $\lambda \in \text{Par}_N$  by adding 1 to  $k$  distinct parts of  $\lambda$ . This means that the skew shape  $\nu/\lambda$  is a vertical strip of size  $k$ .

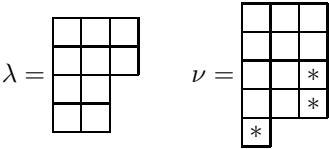
**11.41. Example.** Let  $N = 6$  and  $\lambda = (3, 3, 2, 2)$ . Let  $v$  be the following abacus in  $\text{LAbc}(\lambda + \delta(6))$ :



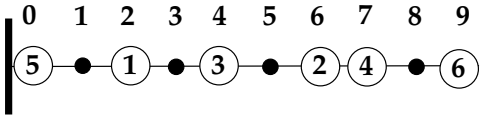
Suppose  $k = 3$  and  $S = \{1, 2, 3\}$ . We move bead 2, then bead 3, then bead 1 one step right on the abacus. No bead collision occurs, and we get the following abacus:



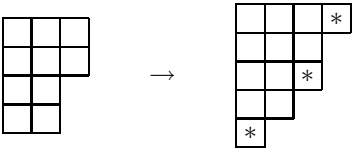
This abacus lies in  $\text{LAbc}(\nu + \delta(6))$ , where  $\nu = (3, 3, 3, 3, 1)$ . Drawing the diagrams, we see that  $\nu$  arises from  $\lambda$  by adding a vertical 3-strip:



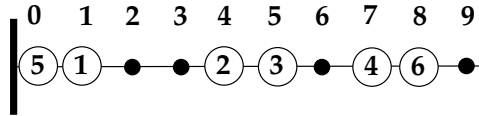
Suppose instead that  $S = \{1, 2, 6\}$ . This time we obtain the abacus



which is in  $\text{LAbc}((4, 3, 3, 2, 1) + \delta(6))$ . Now the partition diagrams look like this:



However, suppose we start with the subset  $S = \{3, 5, 6\}$ . When we move bead 6, then bead 3, then bead 5 on the abacus  $v$ , bead 3 collides with bead 2. We can match the pair  $(v, S)$  to  $(w, T)$ , where  $T = \{2, 5, 6\}$  and  $w$  is this abacus:



Observe that  $\text{sgn}(w) = -\text{sgn}(v)$  and  $\text{wt}(v)x_3x_5x_6 = \text{wt}(w)x_2x_5x_6$ . This example illustrates the cancellation idea used in the proof below.

**11.42. Theorem: Antisymmetric Pieri Rule for  $e_k$ .** For all  $\lambda \in \text{Par}_N$  and all  $k \geq 1$ , the following identity holds in  $K[x_1, \dots, x_N]$ :

$$a_{\lambda+\delta(N)}(x_1, \dots, x_N)e_k(x_1, \dots, x_N) = \sum_{\substack{\beta \in \text{Par}_N: \\ \beta/\lambda \text{ is a vertical } k\text{-strip}}} a_{\beta+\delta(N)}(x_1, \dots, x_N).$$

*Proof.* Let  $X$  be the set of pairs  $(v, S)$  where  $v \in \text{LAbc}(\lambda + \delta(N))$  and  $S$  is a  $k$ -element subset of  $\{1, 2, \dots, N\}$ . Letting  $\text{sgn}(v, S) = \text{sgn}(v)$  and  $\text{wt}(v, S) = \text{wt}(v) \prod_{j \in S} x_j$ , we have

$$a_{\lambda+\delta(N)}e_k = \sum_{z \in X} \text{sgn}(z) \text{wt}(z).$$

Define an involution  $I : X \rightarrow X$  as follows. Given  $(v, S) \in X$ , scan the abacus  $v$  from right to left and move each bead in  $S$  one step to the right. If this can be done with no bead collisions, we obtain an abacus  $v^*$  counted by the sum on the right side of the theorem, such that  $\text{sgn}(v) = \text{sgn}(v^*)$  and  $\text{wt}(v, S) = \text{wt}(v^*)$ . In this case,  $(v, S)$  is a fixed point of  $I$ , and the bead motion rule defines a sign-preserving, weight-preserving bijection between these fixed points and the abaci counted by the right side of the theorem.

Now suppose a bead collision does occur. Then for some  $j \in S$  and some  $k \notin S$ , bead  $k$  lies one step to the right of bead  $j$  in  $v$ . Take  $j$  to be the rightmost bead in  $v$  for which this is true. Let  $I(v, S) = (v', S')$  where  $v'$  is  $v$  with beads  $j$  and  $k$  interchanged, and  $S' = (S \sim \{j\}) \cup \{k\}$ . It is immediately verified that  $\text{sgn}(v', S') = -\text{sgn}(v, S)$ ,  $\text{wt}(v, S) = \text{wt}(v', S')$ , and  $I(v', S') = (v, S)$ . So  $I$  cancels all objects in which a bead collision occurs.  $\square$

## 11.9 Pieri Rule for $h_k$

In the last section, we computed  $a_{\lambda+\delta(N)}e_k$  by using a  $k$ -element *subset* of  $\{1, 2, \dots, N\}$  to move beads on a labeled abacus. Now we will compute  $a_{\lambda+\delta(N)}h_k$  by moving beads based on a  $k$ -element *multiset*. This approach is motivated by the formula

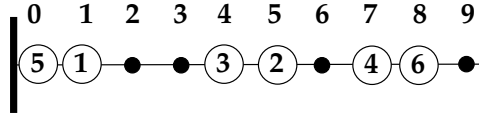
$$h_k(x_1, \dots, x_N) = \sum_{\substack{\text{k-element multisets} \\ M \text{ of } \{1, \dots, N\}}} \prod_{j \in M} x_j,$$

where the factor  $x_j$  is repeated as many times as  $j$  appears in  $M$ .

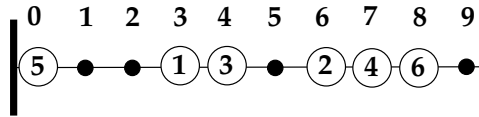
Suppose  $v$  is an abacus counted by  $a_{\lambda+\delta(N)}$ , and  $x_1^{m_1} \cdots x_N^{m_N}$  is a typical term in  $h_k$  (so each  $m_j \geq 0$  and  $m_1 + \cdots + m_N = k$ ). Scan the beads in  $v$  from left to right. Whenever we encounter a bead labeled  $j$ , we move it right, one step at a time, for a total of  $m_j$  positions.

Bead collisions may occur and will lead to object cancellations later. If no collision occurs, we will have a new abacus  $v^* \in \text{LAbc}(\nu + \delta(N))$  with the same sign as  $v$  and weight  $\text{wt}(v^*) = \text{wt}(v)x_1^{m_1} \cdots x_N^{m_N}$ . It follows from the bead motion rule that the shape  $\nu$  arises from  $\lambda$  by adding a *horizontal  $k$ -strip* to  $\lambda$ . Conversely, any abacus indexed by such a shape can be constructed from an abacus indexed by  $\lambda$  by a suitable choice of the bead multiset. These ideas are illustrated in the following example, which should be compared to the example in the preceding section.

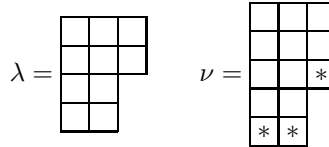
**11.43. Example.** Let  $N = 6$  and  $\lambda = (3, 3, 2, 2)$ . Let  $v$  be the following abacus in  $\text{LAbc}(\lambda + \delta(6))$ :



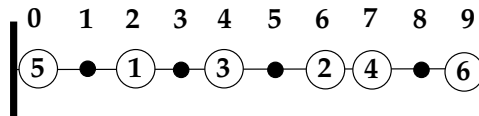
Let  $M$  be the multiset  $[1, 1, 2]$ . It is possible to move bead 1 to the right twice in a row, and then move bead 2 once, without causing any collisions. This produces the following abacus:



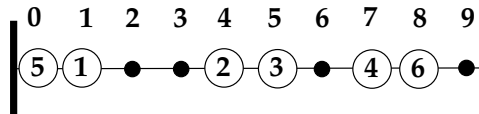
This abacus lies in  $\text{LAbc}(\nu + \delta(6))$ , where  $\nu = (3, 3, 3, 2, 2)$  arises from  $\lambda$  by adding a horizontal 3-strip:



If instead we take  $M = [1, 2, 6]$ , we move bead 1, then bead 2, then bead 6, leading to this abacus in  $\text{LAbc}((4, 3, 3, 2, 1) + \delta(6))$ :



On the other hand, suppose we try to modify  $v$  using the multiset  $M = [1, 2, 3]$ . When scanning  $v$  from left to right, bead 3 moves before bead 2 and collides with bead 2. We match the pair  $(v, M)$  with the pair  $(w, N)$ , where  $N = [1, 2, 2]$  and  $w$  is the following abacus:



Observe that  $\text{sgn}(w) = -\text{sgn}(v)$  and  $\text{wt}(v)x_1x_2x_3 = \text{wt}(w)x_1x_2^2$ . This example illustrates the cancellation idea used in the proof below.

**11.44. Theorem: Antisymmetric Pieri Rule for  $h_k$ .** For all  $\lambda \in \text{Par}_N$  and all  $k \geq 1$ , the following identity holds in  $K[x_1, \dots, x_N]$ :

$$a_{\lambda+\delta(N)}(x_1, \dots, x_N)h_k(x_1, \dots, x_N) = \sum_{\substack{\beta \in \text{Par}_N: \\ \beta/\lambda \text{ is a horizontal } k\text{-strip}}} a_{\beta+\delta(N)}(x_1, \dots, x_N).$$

*Proof.* Let  $X$  be the set of pairs  $(v, M)$  where  $v \in \text{LAbc}(\lambda + \delta(N))$  and  $M = [1^{m_1} 2^{m_2} \dots N^{m_N}]$  is a  $k$ -element multiset. Putting  $\text{sgn}(v, M) = \text{sgn}(v)$  and  $\text{wt}(v, M) = \text{wt}(v) \prod_{j=1}^N x_j^{m_j}$ , we have

$$a_{\lambda+\delta(N)} h_k = \sum_{z \in X} \text{sgn}(z) \text{wt}(z).$$

Define an involution  $I : X \rightarrow X$  as follows. Given  $(v, M) \in X$ , scan the abacus  $v$  from left to right. When bead  $j$  is encountered in the scan, move it  $m_j$  steps right, one step at a time. If all bead motions are completed with no bead collisions, we obtain an abacus  $v^*$  counted by the sum on the right side of the theorem, such that  $\text{sgn}(v) = \text{sgn}(v^*)$  and  $\text{wt}(v, M) = \text{wt}(v^*)$ . In this case,  $(v, M)$  is a fixed point of  $I$ , and the bead motion rule defines a sign-preserving, weight-preserving bijection between these fixed points and the abaci counted by the right side of the theorem.

Now consider the case where a bead collision does occur. Suppose the first collision occurs when bead  $j$  hits a bead  $k$  that is located  $p \leq m_j$  positions to the right of bead  $j$ 's initial position. Define  $I(v, M) = (v', M')$ , where  $v'$  is  $v$  with beads  $j$  and  $k$  interchanged, and  $M'$  is obtained from  $M$  by letting  $j$  occur  $m_j - p \geq 0$  times in  $M'$ , letting  $k$  occur  $m_k + p$  times in  $M'$ , and leaving all other multiplicities the same. One may check that  $\text{sgn}(v', M') = -\text{sgn}(v, M)$ ,  $\text{wt}(v, M) = \text{wt}(v', M')$ , and  $I(v', M') = (v, M)$ . So  $I$  cancels all objects in which a bead collision occurs.  $\square$

## 11.10 Antisymmetric Polynomials and Schur Polynomials

The Pieri rule for computing  $a_{\lambda+\delta(N)} h_k$  closely resembles the rule for computing  $s_{\lambda} h_k$  from §10.12. This resemblance leads to an algebraic proof of a formula expressing Schur polynomials as quotients of antisymmetric polynomials.

**11.45. Theorem: Schur Polynomials and Antisymmetric Polynomials.** For all  $\lambda \in \text{Par}_N$ ,

$$s_{\lambda}(x_1, \dots, x_N) = \frac{a_{\lambda+\delta(N)}(x_1, \dots, x_N)}{a_{\delta(N)}(x_1, \dots, x_N)} = \frac{\det \|x_j^{\lambda_i + N - i}\|_{1 \leq i, j \leq N}}{\det \|x_j^{N - i}\|_{1 \leq i, j \leq N}}.$$

*Proof.* In 10.69, we iterated the Pieri rule

$$s_{\nu}(x_1, \dots, x_N) h_k(x_1, \dots, x_N) = \sum_{\substack{\beta \in \text{Par}_N: \\ \beta/\nu \text{ is a horizontal } k\text{-strip}}} s_{\beta}(x_1, \dots, x_N)$$

to deduce the formula

$$h_{\mu}(x_1, \dots, x_N) = \sum_{\lambda \in \text{Par}_N} K_{\lambda, \mu} s_{\lambda}(x_1, \dots, x_N) \quad (\mu \in \text{Par}). \quad (11.3)$$

Recall that this derivation used semistandard tableaux to encode the sequence of horizontal strips that were added to go from the empty shape to the shape  $\lambda$ . Now, precisely the same idea can be applied to iterate the antisymmetric Pieri rule

$$a_{\nu+\delta(N)}(x_1, \dots, x_N) h_k(x_1, \dots, x_N) = \sum_{\substack{\beta \in \text{Par}_N: \\ \beta/\nu \text{ is a horizontal } k\text{-strip}}} a_{\beta+\delta(N)}(x_1, \dots, x_N).$$

If we start with  $\nu = (0)$  and multiply successively by  $h_{\mu_1}, h_{\mu_2}, \dots$ , we obtain the formula

$$a_{0+\delta(N)}(x_1, \dots, x_N) h_\mu(x_1, \dots, x_N) = \sum_{\lambda \in \text{Par}_N} K_{\lambda, \mu} a_{\lambda+\delta(N)}(x_1, \dots, x_N) \quad (\mu \in \text{Par}). \quad (11.4)$$

Now restrict attention to partitions  $\lambda, \mu \in \text{Par}_N(m)$ . As in 10.72, we can write equations (11.3) in the form  $\mathbf{H} = \mathbf{K}^t \mathbf{S}$ , where  $\mathbf{H} = (h_\mu : \mu \in \text{Par}_N(m))$  and  $\mathbf{S} = (s_\lambda : \lambda \in \text{Par}_N(m))$  are column vectors, and  $\mathbf{K}^t$  is the transpose of the Kostka matrix. Letting  $\mathbf{A} = (a_{\lambda+\delta(N)}/a_{\delta(N)} : \lambda \in \text{Par}_N(m))$ , we can similarly write equations (11.4) in the form  $\mathbf{H} = \mathbf{K}^t \mathbf{A}$ . Finally, since the Kostka matrix is invertible (being unitriangular), we can conclude that

$$\mathbf{A} = (\mathbf{K}^t)^{-1} \mathbf{H} = \mathbf{S}.$$

Equating entries of these vectors gives the desired result.  $\square$

A purely combinatorial proof of the identity  $a_{\lambda+\delta(N)} = s_\lambda a_{\delta(N)}$  will be given in §11.12.

### 11.11 Rim-Hook Tableaux

The connection between Schur polynomials and antisymmetric polynomials lets us deduce the following Pieri rule for calculating the product  $s_\lambda p_k$ .

**11.46. Theorem: Symmetric Pieri Rule for  $p_k$ .** For all  $\lambda \in \text{Par}_N$  and all  $k \geq 1$ , the following identity holds in  $K[x_1, \dots, x_N]$ :

$$s_\lambda(x_1, \dots, x_N) p_k(x_1, \dots, x_N) = \sum_{\substack{\beta \in \text{Par}_N: \\ \beta/\lambda \text{ is a } k\text{-ribbon } R}} \text{sgn}(R) s_\beta(x_1, \dots, x_N).$$

*Proof.* Start with the identity

$$a_{\lambda+\delta(N)}(x_1, \dots, x_N) p_k(x_1, \dots, x_N) = \sum_{\substack{\beta \in \text{Par}_N: \\ \beta/\lambda \text{ is a } k\text{-ribbon } R}} \text{sgn}(R) a_{\beta+\delta(N)}(x_1, \dots, x_N)$$

(proved in 11.39), divide both sides by  $a_{\delta(N)}$ , and use 11.45.  $\square$

**11.47. Example.** Suppose we multiply  $s_{(0)} = 1$  by  $p_4$  using the Pieri rule. The result is a signed sum of Schur polynomials indexed by 4-ribbons:

$$p_4 = s_{(0)} p_4 = s_{(4)} - s_{(3,1)} + s_{(2,1,1)} - s_{(1,1,1,1)}.$$

To expand  $p_{(4,3)}$  into Schur polynomials, first multiply both sides of the previous equation by  $p_3$ :

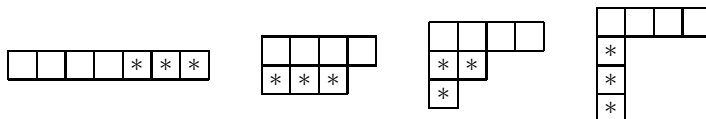
$$p_{(4,3)} = p_4 p_3 = s_{(4)} p_3 - s_{(3,1)} p_3 + s_{(2,1,1)} p_3 - s_{(1,1,1,1)} p_3.$$

Now use the Pieri rule on each term on the right side. This leads to the diagrams shown in Figure 11.3. Taking signs into account, this leads to the formula

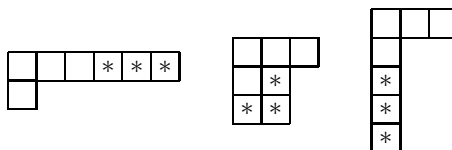
$$\begin{aligned} p_{(4,3)} &= s_{(7)} + s_{(4,3)} - s_{(4,2,1)} + s_{(4,1,1,1)} \\ &\quad - s_{(6,1)} + s_{(3,2,2)} - s_{(3,1,1,1,1)} \\ &\quad + s_{(5,1,1)} - s_{(3,3,1)} + s_{(2,1,1,1,1,1)} \\ &\quad - s_{(4,1,1,1,1)} + s_{(3,2,1,1)} - s_{(2,2,2,1)} - s_{(1,1,1,1,1,1,1)}. \end{aligned}$$

Here we are assuming  $N$  (the number of variables) is at least 7.

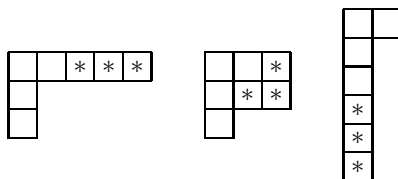
Shapes for  $s_{(4)}p_3$ :



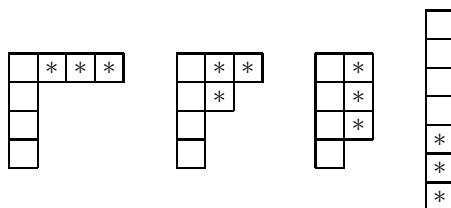
Shapes for  $-s_{(3,1)}p_3$ :



Shapes for  $s_{(2,1,1)}p_3$ :



Shapes for  $-s_{(1,1,1,1)}p_3$ :



**FIGURE 11.3**

Adding  $k$ -ribbons to compute  $s_{\lambda}p_k$ .

Just as we used semistandard tableaux to encode successive additions of horizontal strips, we can use the following notion of a *rim-hook tableau* to encode successive additions of signed ribbons.

**11.48. Definition: Rim-Hook Tableaux.** Given a partition  $\lambda$  and a sequence  $\alpha \in \mathbb{N}^s$ , a *rim-hook tableau of shape  $\lambda$  and content  $\alpha$*  is a sequence  $T$  of partitions

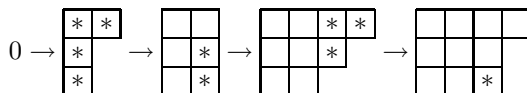
$$(0) = \nu^0 \subseteq \nu^1 \subseteq \nu^2 \subseteq \cdots \subseteq \nu^s = \lambda$$

such that  $\nu^i/\nu^{i-1}$  is an  $\alpha_i$ -ribbon for  $1 \leq i \leq s$ . We represent this tableau pictorially by drawing the diagram of  $\lambda$  and entering the number  $i$  in each cell of the ribbon  $\nu^i/\nu^{i-1}$ . The *sign* of the rim-hook tableau  $T$  is the product of the signs of the ribbons  $\nu^i/\nu^{i-1}$ . (Recall that the sign of a ribbon occupying  $r$  rows is  $(-1)^{r-1}$ .) Let  $\text{RHT}(\lambda, \alpha)$  be the set of all rim-hook tableaux of shape  $\lambda$  and content  $\alpha$ . Finally, define the integer

$$\chi_\alpha^\lambda = \sum_{T \in \text{RHT}(\lambda, \alpha)} \text{sgn}(T).$$

Rim-hook tableaux of skew shape  $\lambda/\mu$  are defined analogously; now we require that  $\nu^0 = \mu$ , so that the cells of  $\mu$  do not get filled with ribbons. The set  $\text{RHT}(\lambda/\mu, \alpha)$  and the integer  $\chi_\alpha^{\lambda/\mu}$  are defined as above.

**11.49. Example.** Suppose we expand the product  $p_4 p_2 p_3 p_1$  into a sum of Schur polynomials. We can do this by applying the Pieri rule four times, starting with the empty shape. Each application of the Pieri rule will add a new border ribbon to the shape. The lengths of the ribbons are given by the content vector  $\alpha = (4, 2, 3, 1)$ . Here is one possible sequence of ribbon additions:



This sequence of shapes defines a rim-hook tableau

$$T = ((0), (2, 1, 1), (2, 2, 2), (4, 3, 2), (4, 3, 3))$$

which can be visualized using the following diagram:

$$T = \begin{array}{|c|c|c|c|} \hline 1 & 1 & 3 & 3 \\ \hline 1 & 2 & 3 & \\ \hline 1 & 2 & 4 & \\ \hline \end{array}.$$

Note that the ribbons we added have signs  $+1$ ,  $-1$ ,  $-1$ , and  $+1$ , so  $\text{sgn}(T) = +1$ . This particular choice of ribbon additions will therefore produce a term  $+s_{(4,3,3)}$  in the Schur expansion of  $p_{(4,2,3,1)}$ .

Now suppose we want to know the coefficient of  $s_{(4,3,3)}$  in the Schur expansion of  $p_4 p_2 p_3 p_1$ . The preceding discussion shows that we will obtain a term  $\pm s_{(4,3,3)}$  for every rim-hook tableau of shape  $(4, 3, 3)$  and content  $(4, 2, 3, 1)$ , where the sign of the term is the sign of the tableau. To find the desired coefficient, we must enumerate all the objects in  $\text{RHT}((4, 3, 3), (4, 2, 3, 1))$ . In addition to the tableau  $T$  displayed above, we find the following tableaux:

$$\begin{array}{|c|c|c|c|} \hline 1 & 1 & 3 & 4 \\ \hline 1 & 2 & 3 & \\ \hline 1 & 2 & 3 & \\ \hline \end{array} \quad \begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & 2 \\ \hline 1 & 3 & 3 & \\ \hline 1 & 3 & 4 & \\ \hline \end{array} \quad \begin{array}{|c|c|c|c|} \hline 1 & 1 & 1 & 4 \\ \hline 1 & 2 & 2 & \\ \hline 3 & 3 & 3 & \\ \hline \end{array} \quad \begin{array}{|c|c|c|c|} \hline 1 & 1 & 1 & 1 \\ \hline 2 & 3 & 3 & \\ \hline 2 & 3 & 4 & \\ \hline \end{array}$$

The signs of the new tableaux are  $-1$ ,  $-1$ ,  $-1$ , and  $+1$ , so the coefficient is  $+1 - 1 - 1 - 1 + 1 = -1$ .

The calculations in the preceding example generalize to give the following rule for expanding power-sum polynomials into sums of Schur polynomials.

**11.50. Theorem: Schur Expansion of Power-Sum Polynomials.** For all vectors  $\alpha \in \mathbb{N}^t$  and all  $N \geq 1$ ,

$$p_\alpha(x_1, \dots, x_N) = \sum_{\lambda \in \text{Par}_N} \chi_\alpha^\lambda s_\lambda(x_1, \dots, x_N).$$

*Proof.* By iteration of the Pieri rule, the coefficient of  $s_\lambda$  in  $p_\alpha = s_{(0)} p_{\alpha_1} \cdots p_{\alpha_t}$  is the signed sum of all sequences of partitions

$$0 = \nu^0 \subseteq \nu^1 \subseteq \nu^2 \subseteq \cdots \subseteq \nu^t = \lambda$$

such that the skew shape  $\nu^i / \nu^{i-1}$  is an  $\alpha_i$ -ribbon for all  $i$ . By the very definition of rim-hook tableaux, this sum is precisely  $\chi_\alpha^\lambda$ .  $\square$

**11.51. Theorem: Symmetry of  $\chi_\alpha^\lambda$ .** If  $\alpha$  and  $\beta$  are compositions with  $\text{sort}(\alpha) = \text{sort}(\beta)$ , then  $\chi_\alpha^\lambda = \chi_\beta^\lambda$  for all partitions  $\lambda$ .

*Proof.* The hypothesis implies that the sequence  $\alpha = (\alpha_1, \alpha_2, \dots)$  can be rearranged to the sequence  $\beta = (\beta_1, \beta_2, \dots)$ . It follows from this that  $p_\alpha = \prod p_{\alpha_i} = \prod p_{\beta_i} = p_\beta$ , since multiplication of polynomials is commutative. Let  $k = \sum_i \alpha_i$  and take  $N \geq k$ . Two applications of the previous theorem give

$$\sum_{\lambda \in \text{Par}(k)} \chi_\alpha^\lambda s_\lambda = p_\alpha = p_\beta = \sum_{\lambda \in \text{Par}(k)} \chi_\beta^\lambda s_\lambda.$$

By linear independence of the Schur polynomials  $\{s_\lambda(x_1, \dots, x_N) : \lambda \in \text{Par}(k)\}$ , we conclude that  $\chi_\alpha^\lambda = \chi_\beta^\lambda$  for all  $\lambda$ .  $\square$

**11.52. Remark.** The last theorem and corollary extend to skew shapes as follows. If  $\mu$  is a partition, then

$$s_\mu p_\alpha = \sum_{\substack{\lambda \in \text{Par}_N: \\ \mu \subseteq \lambda}} \chi_\alpha^{\lambda/\mu} s_\lambda.$$

Furthermore, if  $\text{sort}(\alpha) = \text{sort}(\beta)$  then  $\chi_\alpha^{\lambda/\mu} = \chi_\beta^{\lambda/\mu}$ . The proof is the same as before, replacing (0) by  $\mu$  throughout.

We have just seen how to expand power-sum symmetric polynomials into sums of Schur polynomials. Conversely, it is possible to express Schur polynomials in terms of the  $p_\mu$ 's. We can use the Hall scalar product from §10.26 to derive this expansion from the previous one.

**11.53. Theorem: Power-Sum Expansion of Schur Polynomials.** For  $N \geq k$  and all  $\lambda \in \text{Par}(k)$ ,

$$s_\lambda(x_1, \dots, x_N) = \sum_{\mu \in \text{Par}(k)} \frac{\chi_\mu^\lambda}{z_\mu} p_\mu(x_1, \dots, x_N).$$

*Proof.* For all  $\mu \in \text{Par}(k)$ , we know that  $p_\mu = \sum_{\nu \in \text{Par}(k)} \chi_\mu^\nu s_\nu$ . Therefore, for a given partition  $\lambda \in \text{Par}(k)$ ,

$$\langle p_\mu, s_\lambda \rangle = \sum_{\nu \in \text{Par}(k)} \chi_\mu^\nu \langle s_\nu, s_\lambda \rangle = \chi_\mu^\lambda$$



since the Schur polynomials are orthonormal relative to the Hall scalar product. Now, since the  $p_\mu$ 's form a basis of  $\Lambda_N^k$ , we know there exist scalars  $c_\nu \in \mathbb{Q}$  with  $s_\lambda = \sum_\nu c_\nu p_\nu$ . To find a given coefficient  $c_\mu$ , we compute

$$\chi_\mu^\lambda = \langle p_\mu, s_\lambda \rangle = \sum_\nu c_\nu \langle p_\mu, p_\nu \rangle = c_\mu z_\mu,$$

where the last equality follows by definition of the Hall scalar product. We see that  $c_\mu = \chi_\mu^\lambda / z_\mu$ , as desired.  $\square$

## 11.12 Abaci and Tableaux

This section contains a combinatorial proof of the identity

$$a_{\delta(N)}(x_1, \dots, x_N) s_\lambda(x_1, \dots, x_N) = a_{\lambda+\delta(N)}(x_1, \dots, x_N),$$

which we proved algebraically in §11.10.

Let  $X$  be the set of pairs  $(v, T)$ , where  $v$  is a justified labeled abacus with  $N$  beads and  $T$  is a semistandard tableau using letters in  $\{1, 2, \dots, N\}$ . It will be convenient to use the following non-standard total ordering on this alphabet that depends on  $v$ :  $i <_v j$  iff bead  $i$  is to the right of bead  $j$  on the abacus  $v$ . Equivalently, we can describe the total order by writing

$$v_{N-1} <_v v_{N-2} <_v \dots <_v v_1 <_v v_0.$$

Here are two examples of objects in  $X$  when  $N = 7$  and  $\lambda = (7, 7, 5, 3, 2)$ :

$$(v^1, T^1) = \left( 7654321000 \dots, \begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ \hline 3 & 3 & 5 & 5 & 5 & & \\ \hline 6 & 6 & 6 & & & & \\ \hline 7 & 7 & & & & & \\ \hline \end{array} \right);$$

$$(v^2, T^2) = \left( 2451763000 \dots, \begin{array}{|c|c|c|c|c|c|c|} \hline 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ \hline 6 & 6 & 6 & 6 & 6 & 6 & 6 \\ \hline 7 & 7 & 5 & 5 & 5 & & \\ \hline 4 & 4 & 4 & & & & \\ \hline 2 & 2 & & & & & \\ \hline \end{array} \right).$$

Note that we can pass from the first tableau (which is semistandard under the usual ordering) to the second tableau (which is semistandard relative to one of the non-standard orderings) by applying the permutation  $7 \mapsto 2, 6 \mapsto 4, \dots$  to each entry in the first tableau. It follows that the generating function for the set  $\text{SSYT}_N(\lambda)$  relative to one of the orderings  $<_v$  can be obtained from the usual generating function for semistandard tableaux (namely  $s_\lambda(x_1, \dots, x_N)$ ) by applying the permutation  $x_7 \mapsto x_2, x_6 \mapsto x_4, \dots$ . Since Schur polynomials are symmetric, the answer is still  $s_\lambda(x_1, \dots, x_N)$ . By the product rule for weighted sets, we conclude that

$$\sum_{(v, T) \in X} \text{sgn}(v) \text{wt}(v) \text{wt}(T) = a_{\delta(N)}(x_1, \dots, x_N) s_\lambda(x_1, \dots, x_N).$$

On the other hand, the generating function for the set  $Y$  of  $N$ -bead labeled abaci with beads

in positions  $\lambda + \delta(N)$  is  $a_{\lambda + \delta(N)}(x_1, \dots, x_N)$ . So it suffices to define a sign-reversing, weight-preserving involution  $I : X \rightarrow X$  where the fixed point set of  $I$  corresponds bijectively to  $Y$ .

The main idea is that the tableau  $T$  encodes a sequence of bead motions on the abacus  $v$ . If performing these movements causes a bead collision, then  $(v, T)$  will cancel with some other object in  $X$ . Otherwise, the abacus obtained from  $v$  by the bead motions will be one of the objects in  $Y$ .

A tableau  $T$  specifies bead motions as follows. Define the *word of  $T$*  to be the word  $w(T) = w_1 w_2 \cdots w_n$  (where  $n = |\lambda|$ ) obtained by concatenating the rows of  $T$  from bottom to top. For example, the object  $(v^2, T^2)$  shown above has

$$w(T^2) = 224447755566666663333333.$$

Now, given  $(v, T) \in X$ , scan the symbols in  $w(T)$  from right to left. When a symbol  $j$  is encountered, move the bead labeled  $j$  in  $v$  one step to the right.

Let us first determine which objects  $(v, T)$  have no bead collisions. Suppose  $v = v_0 \dots v_{N-1} 00 \dots$ . Let  $i$  be the last entry in the top row of  $T$ , which is the rightmost letter in  $w(T)$ . We must first move bead  $i$  one step to the right. This move will already cause a collision (since  $v$  is justified) unless  $i = v_{N-1}$ . Since  $v_{N-1}$  is the smallest letter relative to  $<_v$  and  $T$  is semistandard,  $i = v_{N-1}$  iff all entries in the top row of  $T$  are equal to  $v_{N-1}$ . In this situation, we will move the rightmost bead  $v_{N-1}$  to the right  $\lambda_1$  positions with no collisions.

Now we repeat the argument on the second row of  $T$ . The rightmost entry  $j$  in this row cannot be  $v_{N-1}$  (otherwise we would not have a strict increase in every column). The only way to avoid an immediate bead collision is when  $j = v_{N-2}$ , in which case all entries in the second row must equal  $v_{N-2}$ . In this situation, bead  $v_{N-2}$  will move to the right  $\lambda_2$  positions with no collisions.

Continuing similarly, we see that  $(v, T)$  will have no collisions iff for all  $k$ , the  $k$ th row of  $T$  consists of  $\lambda_k$  copies of the  $k$ th smallest letter  $v_{N-k}$ . Moving the beads on  $v$  according to  $T$  has the effect of unjustifying  $v$  to an abacus  $v^* \in Y = \text{LAbc}(\lambda + \delta(N))$ . Defining  $I(v, T) = (v, T)$  in this case, we therefore have specified a bijection between the fixed points of  $I$  and  $Y$ . For example,

$$(v, T) = \left( 2451763000 \cdots, \begin{array}{|c|c|c|c|c|c|c|} \hline 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ \hline 6 & 6 & 6 & 6 & 6 & 6 & 6 \\ \hline 7 & 7 & 7 & 7 & 7 & & \\ \hline 1 & 1 & 1 & & & & \\ \hline 5 & 5 & & & & & \\ \hline \end{array} \right) \mapsto v^* = 24005010070063000 \cdots.$$

The map  $(v, T) \mapsto v^*$  preserves signs and weights.

To complete the proof, we describe a cancellation mechanism to pair off objects  $(v, T)$  in which bead collisions do occur. Suppose the first bead collision for  $(v, T)$  occurs when some bead  $i$  moves to the right one step and bumps into bead  $j$ . Note that  $i >_v j$ , and  $i, j$  must be two adjacent letters in the total ordering  $>_v$ . Define  $(v', T') = I(v, T)$  as follows. We obtain  $v'$  from  $v$  by interchanging the adjacent beads  $i$  and  $j$ , so that  $\text{sgn}(v') = -\text{sgn}(v)$ ,  $\text{wt}(v')x_j = \text{wt}(v)x_i$ , and  $<_{v'}$  agrees with  $<_v$  except that now  $i <_{v'} j$ .

We obtain  $T'$  from  $T$  by modifying the occurrences of  $i$  and  $j$  in  $w(T)$  by a procedure similar to the one used in §10.6. By the argument used to determine the fixed points of  $I$ , we know that the occurrence of  $i$  in  $T$  that caused the bead collision is the rightmost entry in some row of  $T$ , say the  $k$ th row; furthermore, for  $1 \leq l < k$ , row  $l$  consists of  $\lambda_l$  copies of  $v_{N-l}$ . Now  $i >_v v_{N-k}$  (or this entry of  $T$  would not cause a collision), and so  $j \geq_v v_{N-k}$ . This means that no entry in the first  $k-1$  rows of  $T$  equals  $i$  or  $j$ , so these rows can be ignored in the following discussion.

We now describe how to change  $T$  into  $T'$ . Whenever  $j$  occurs directly above  $i$  in  $T$  (call these occurrences *matched pairs*), interchange these two symbols. Some rows of  $T$  will contain unmatched  $i$ 's and  $j$ 's, in which  $a \geq 0$  copies of  $j$  are followed by  $b \geq 0$  copies of  $i$ . In particular, row  $k$  will have  $a \geq 0$  and  $b > 0$ , since the  $i$  at the end of the row cannot be matched with a  $j$  above it. In row  $k$ , replace the unmatched symbols  $j^a i^b$  by  $j^{a+1} i^{b-1}$ . Then, in all rows containing unmatched  $i$ 's and  $j$ 's (including the new row  $k$ ), replace the unmatched symbols  $j^a i^b$  by  $i^b j^a$ . The following assertions can now be checked:  $T'$  is a semistandard tableau relative to  $<_{v'}$ ;  $T'$  has one fewer  $i$  and one more  $j$  than  $T$  does;  $\text{wt}(T')x_i = \text{wt}(T)x_j$ ;  $\text{wt}(v', T') = \text{wt}(v, T)$ ;  $\text{sgn}(v', T') = -\text{sgn}(v)$ ; the last symbol in row  $k$  of  $T'$  is an unmatched  $j$ ; this unmatched  $j$  will cause the first bead collision when  $T'$  is used to move the beads on  $v'$ ; and  $I(v', T') = (v, T)$ .

**11.54. Example.** Consider the object

$$(v, T) = \left( 2451763000 \cdots, \begin{array}{|c|c|c|c|c|c|c|} \hline 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ \hline 6 & 6 & 6 & 6 & 6 & 6 & 6 \\ \hline 7 & 7 & 5 & 5 & 5 & & \\ \hline 4 & 4 & 4 & & & & \\ \hline 2 & 2 & & & & & \\ \hline \end{array} \right).$$

Processing the first two rows of  $T$ , we move bead 3 right seven positions, then move bead 6 right 7 positions with no collisions. But in row 3, the rightmost symbol  $i = 5$  causes a collision with bead  $j = 1$ . There are no matched pairs of 5's and 1's in this tableau, so we first change the 555 in row 3 to 155, and then change this string to 551 to preserve semistandardness under the new ordering. We have

$$I(v, T) = \left( 2415763000 \cdots, \begin{array}{|c|c|c|c|c|c|c|} \hline 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ \hline 6 & 6 & 6 & 6 & 6 & 6 & 6 \\ \hline 7 & 7 & 5 & 5 & 1 & & \\ \hline 4 & 4 & 4 & & & & \\ \hline 2 & 2 & & & & & \\ \hline \end{array} \right).$$

If we apply  $I$  to this object, bead 1 bumps into bead 5, and we find that  $I(I(v, T)) = (v, T)$ .

**11.55. Example.** Consider the object

$$(v, T) = \left( 2451763000 \cdots, \begin{array}{|c|c|c|c|c|c|c|} \hline 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ \hline 6 & 6 & 6 & 7 & 7 & 7 & 7 \\ \hline 7 & 7 & 5 & 5 & 5 & & \\ \hline 4 & 4 & 4 & & & & \\ \hline 2 & 2 & & & & & \\ \hline \end{array} \right).$$

Now the first collision occurs when bead  $i = 7$  bumps into bead  $j = 6$  because of the 7 at the end of the second row of  $T$ . The first two 6's in that row are matched with 7's below, so the unmatched  $i$ 's and  $j$ 's in row 2 are 67777. We replace this string first by 66777, and then by 77766. Interchanging the matched 6's and 7's leads to

$$I(v, T) = \left( 2451673000 \cdots, \begin{array}{|c|c|c|c|c|c|c|} \hline 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ \hline 7 & 7 & 7 & 7 & 7 & 6 & 6 \\ \hline 6 & 6 & 5 & 5 & 5 & & \\ \hline 4 & 4 & 4 & & & & \\ \hline 2 & 2 & & & & & \\ \hline \end{array} \right).$$

**11.56. Example.** The reader may check that

$$I \left( 76543210 \cdots, \begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 2 & 2 & 3 & 4 \\ \hline 2 & 3 & 3 & 3 & 4 & 4 & 7 \\ \hline 3 & 4 & 5 & 5 & 5 & & \\ \hline \end{array} \right) = \left( 76534210 \cdots, \begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 2 & 2 & 4 & 3 \\ \hline 2 & 4 & 4 & 3 & 3 & 3 & 7 \\ \hline 3 & 3 & 5 & 5 & 5 & & \\ \hline \end{array} \right).$$

### 11.13 Skew Schur Polynomials

In the remainder of this chapter, we will develop further combinatorial properties of skew Schur polynomials. Recall the definition from 10.14: for every skew shape  $\lambda/\mu$ ,

$$s_{\lambda/\mu}(x_1, \dots, x_N) = \sum_{T \in \text{SSYT}_N(\lambda/\mu)} x^{c(T)}.$$

In 10.35, we proved that skew Schur polynomials are symmetric. More precisely, we have the expansion in the monomial basis:

$$s_{\lambda/\mu}(x_1, \dots, x_N) = \sum_{\nu \in \text{Par}_N} K_{\lambda/\mu, \nu} m_\nu(x_1, \dots, x_N),$$

where  $K_{\lambda/\mu, \nu}$  is the number of semistandard tableaux of shape  $\lambda/\mu$  and content  $\nu$ . Our current goal is to find combinatorial formulas for the expansion of skew Schur polynomials relative to some other bases for  $\Lambda$ . We begin by proving an algebraic fact involving the Hall scalar product.

**11.57. Theorem: Skew Schur Polynomials and the Hall Scalar Product.** Suppose  $\lambda, \mu \in \text{Par}$ ,  $k = |\lambda| - |\mu|$ ,  $N \geq |\lambda|$ , and  $f \in \Lambda_N^k$ . Then  $\langle s_{\lambda/\mu}, f \rangle = \langle s_\lambda, s_\mu f \rangle$ .

*Proof.* We first prove the result for  $f = h_\nu$ , where  $\nu \in \text{Par}(k)$ . On one hand, we have the expansion

$$s_{\lambda/\mu} = \sum_{\rho \in \text{Par}(k)} K_{\lambda/\mu, \rho} m_\rho.$$

Taking the scalar product of both sides with  $h_\nu$  gives  $\langle s_{\lambda/\mu}, h_\nu \rangle = K_{\lambda/\mu, \nu}$  (see 10.132).

On the other hand, the Pieri rule shows that

$$s_\mu h_\nu = \sum_{\rho} K_{\rho/\mu, \nu} s_\rho$$

(see 10.71). Taking the scalar product with  $s_\lambda$  gives  $\langle s_\lambda, s_\mu h_\nu \rangle = K_{\lambda/\mu, \nu}$ . Thus the result holds for every  $f$  in the complete homogeneous basis.

The general case now follows by linearity: given any  $f \in \Lambda_N^k$ , write  $f = \sum_\nu c_\nu h_\nu$  for certain scalars  $c_\nu \in K$ . Then compute

$$\begin{aligned} \langle s_{\lambda/\mu}, f \rangle &= \left\langle s_{\lambda/\mu}, \sum_\nu c_\nu h_\nu \right\rangle = \sum_\nu c_\nu \langle s_{\lambda/\mu}, h_\nu \rangle \\ &= \sum_\nu c_\nu \langle s_\lambda, s_\mu h_\nu \rangle = \left\langle s_\lambda, \sum_\nu c_\nu s_\mu h_\nu \right\rangle = \langle s_\lambda, s_\mu f \rangle. \quad \square \end{aligned}$$

We can use 11.57 to expand skew Schur polynomials in terms of power-sum symmetric polynomials.

**11.58. Theorem: Power-Sum Expansion of Skew Schur Polynomials.** Suppose  $\mu \subseteq \lambda$  are partitions with  $k = |\lambda| - |\mu|$ . For all  $N \geq |\lambda|$ ,

$$s_{\lambda/\mu}(x_1, \dots, x_N) = \sum_{\nu \in \text{Par}(k)} \frac{\chi_\nu^{\lambda/\mu}}{z_\nu} p_\nu(x_1, \dots, x_N).$$

*Proof.* We imitate the proof of 11.53. Start with the expansion

$$s_\mu p_\nu = \sum_{\lambda} \chi_\nu^{\lambda/\mu} s_\lambda.$$

Now take the scalar product of both sides with a given partition  $\lambda$ :

$$\langle s_\lambda, s_\mu p_\nu \rangle = \chi_\nu^{\lambda/\mu}.$$

We know the symmetric polynomial  $s_{\lambda/\mu}$  has some expansion in the power-sum basis, say  $s_{\lambda/\mu} = \sum_{\nu} a_\nu p_\nu$  for some  $a_\nu \in K$ . To find a particular  $a_\nu$ , take the scalar product with  $p_\nu/z_\nu$  to get

$$a_\nu = \langle s_{\lambda/\mu}, p_\nu/z_\nu \rangle = \langle s_\lambda, s_\mu p_\nu/z_\nu \rangle = \langle s_\lambda, s_\mu p_\nu \rangle / z_\nu = \chi_\nu^{\lambda/\mu} / z_\nu. \quad \square$$

We also deduce the effect of the involution  $\omega$  on skew Schur polynomials.

**11.59. Theorem: Action of  $\omega$  on Skew Schur Polynomials.** For all partitions  $\mu \subseteq \lambda$  and all  $N \geq |\lambda|$ ,

$$\omega(s_{\lambda/\mu}(x_1, \dots, x_N)) = s_{\lambda'/\mu'}(x_1, \dots, x_N).$$

*Proof.* We already know that the involution  $\omega$  is a ring homomorphism and isometry sending every  $s_\alpha$  to  $s_{\alpha'}$ . For each partition  $\nu$  of size  $|\lambda| - |\mu|$ , we can therefore write:

$$\begin{aligned} \langle \omega(s_{\lambda/\mu}), s_\nu \rangle &= \langle \omega^2(s_{\lambda/\mu}), \omega(s_\nu) \rangle = \langle s_{\lambda/\mu}, s_{\nu'} \rangle = \langle s_\lambda, s_\mu s_{\nu'} \rangle \\ &= \langle \omega(s_\lambda), \omega(s_\mu s_{\nu'}) \rangle = \langle s_{\lambda'}, s_{\mu'} s_\nu \rangle = \langle s_{\lambda'/\mu'}, s_\nu \rangle. \end{aligned}$$

Thus  $\omega(s_{\lambda/\mu})$  and  $s_{\lambda'/\mu'}$  have the same expansion in the Schur basis and are therefore equal.  $\square$

## 11.14 Jacobi-Trudi Formulas

Our next goal is to obtain formulas expressing skew Schur polynomials as determinants involving the complete symmetric polynomials  $h_k$  or the elementary symmetric polynomials  $e_k$ . To derive these results, we need a new combinatorial construction relating tableaux to collections of non-intersecting lattice paths.

We begin by interpreting  $h_k(x_1, \dots, x_N)$  in terms of lattice paths. Fix an integer  $a$  and consider the set  $S$  of lattice paths from  $(a, 1)$  to  $(a+k, N)$  that take unit steps up (u) and east (e). We can encode a path  $p$  in this set by listing the  $y$ -coordinates of the successive east steps of  $p$ . For example, the path  $eeuuueee$  corresponds to the sequence  $1, 1, 3, 4, 4$ . This gives a bijection from  $S$  to the set of weakly increasing sequences  $1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq N$ . Let us weight the path corresponding to this sequence by  $x_{i_1} x_{i_2} \dots x_{i_k}$ . Comparing to the definition of  $h_k$ , we see that

$$h_k(x_1, \dots, x_N) = \sum_{p \in S} \text{wt}(p).$$

This formula holds for all integers  $k$  (possibly negative), provided we take  $h_0 = 1$  and  $h_k = 0$  for negative  $k$ .

Now let  $\lambda$  be a partition with  $n \leq N$  parts, and let  $\mu \subseteq \lambda$ . Let  $X$  be the set of fillings of the skew shape  $\lambda/\mu$  using letters in  $\{1, 2, \dots, N\}$  such that each row weakly increases. Let

$Y$  be the set of sequences  $P = (p_1, \dots, p_n)$  where  $p_i$  is a lattice path from  $(n - i + \mu_i, 1)$  to  $(n - i + \lambda_i, N)$ . Let  $\text{wt}(P) = \text{wt}(p_1) \cdots \text{wt}(p_n)$  record the  $y$ -coordinates of all the east steps of the paths in  $P$ . As explained above, we can encode each row  $i$  of a filling  $U \in X$  as a lattice path  $p_i$  from  $(a, 1)$  to  $(a + \lambda_i - \mu_i, N)$ , where  $a = n - i + \mu_i$ . The association  $U \mapsto (p_1, \dots, p_n)$  defines a weight-preserving bijection  $f : X \rightarrow Y$ . Some examples are shown in Figure 11.4.

We say that two lattice paths *intersect* iff they share a common edge or vertex. Let  $Y'$  be the set of  $P \in Y$  such that no two paths in  $P$  intersect. Inspection of Figure 11.4 suggests that  $f$  restricts to a weight-preserving bijection from  $\text{SSYT}_N(\lambda/\mu)$  to  $Y'$ . To see why this holds, consider consecutive entries  $U(i, j) = a$  and  $U(i + 1, j) = b$  in column  $j$  of a filling  $U \in X$ . In  $f(U)$ , path  $p_i$  has an east step from  $(n - i + \mu_i + (j - \mu_i) - 1, a) = (n + j - i - 1, a)$  to  $(n + j - i, a)$ , whereas  $p_{i+1}$  has an east step from  $(n + j - i - 2, b)$  to  $(n + j - i - 1, b)$ . Suppose  $a \geq b$ . Since the beginning of  $p_i$  goes from  $(n - i + \mu_i, 1)$  to  $(n + j - i - 1, a)$ , there is no way for  $p_{i+1}$  (which starts to the left of  $p_i$ ) to reach the point  $(n + j - i - 1, b)$  without intersecting  $p_i$ . Conversely, suppose two paths intersect. Then there must exist  $i$  such that  $p_i$  and  $p_{i+1}$  intersect. The earliest intersection of these paths must occur when  $p_{i+1}$  “bumps into”  $p_i$  by taking an east step ending at some point  $(n + j - i - 1, b)$ . One may now check that there must exist an east step in  $p_i$  starting at  $(n + j - i - 1, a)$  for some  $a \geq b$ , which shows that  $U(i, j) \geq U(i + 1, j)$  in the filling  $U$ .

Now we are ready to prove the Jacobi-Trudi formulas. The idea is to introduce a large collection of signed, weighted sequences of paths that model the terms of a determinant. Cancellations will remove all sequences of intersecting paths, leaving only the objects in  $Y'$ , which correspond to semistandard skew tableaux.

**11.60. Theorem: Jacobi-Trudi Formula.** Suppose  $\lambda$  is a partition with  $n \leq N$  parts, and  $\mu \subseteq \lambda$ . Then

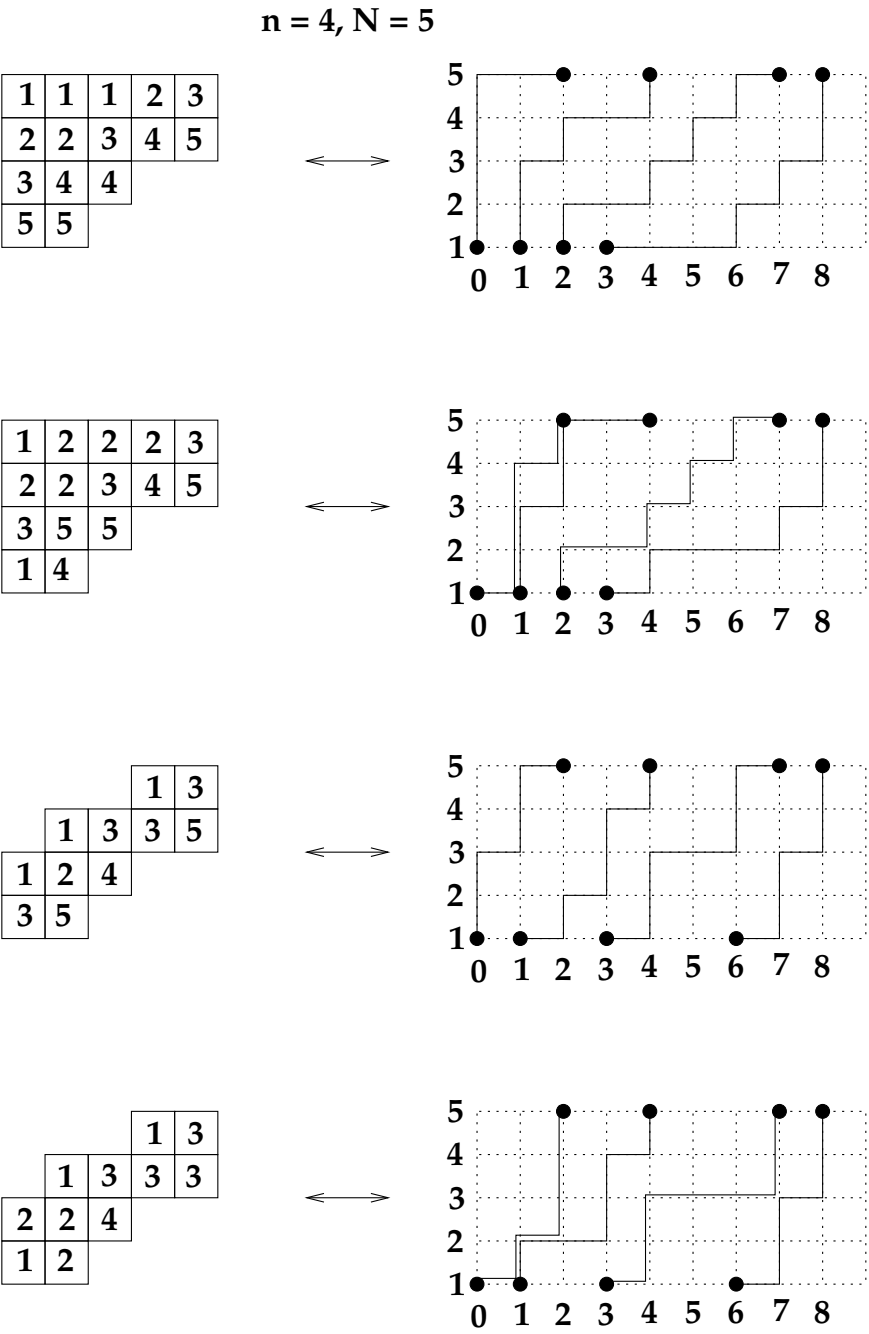
$$s_{\lambda/\mu}(x_1, \dots, x_N) = \det ||h_{\lambda_i - \mu_j + j - i}(x_1, \dots, x_N)||_{1 \leq i, j \leq n}.$$

*Proof.* By the definition of a determinant (see 9.37), the right side of the desired formula can be written

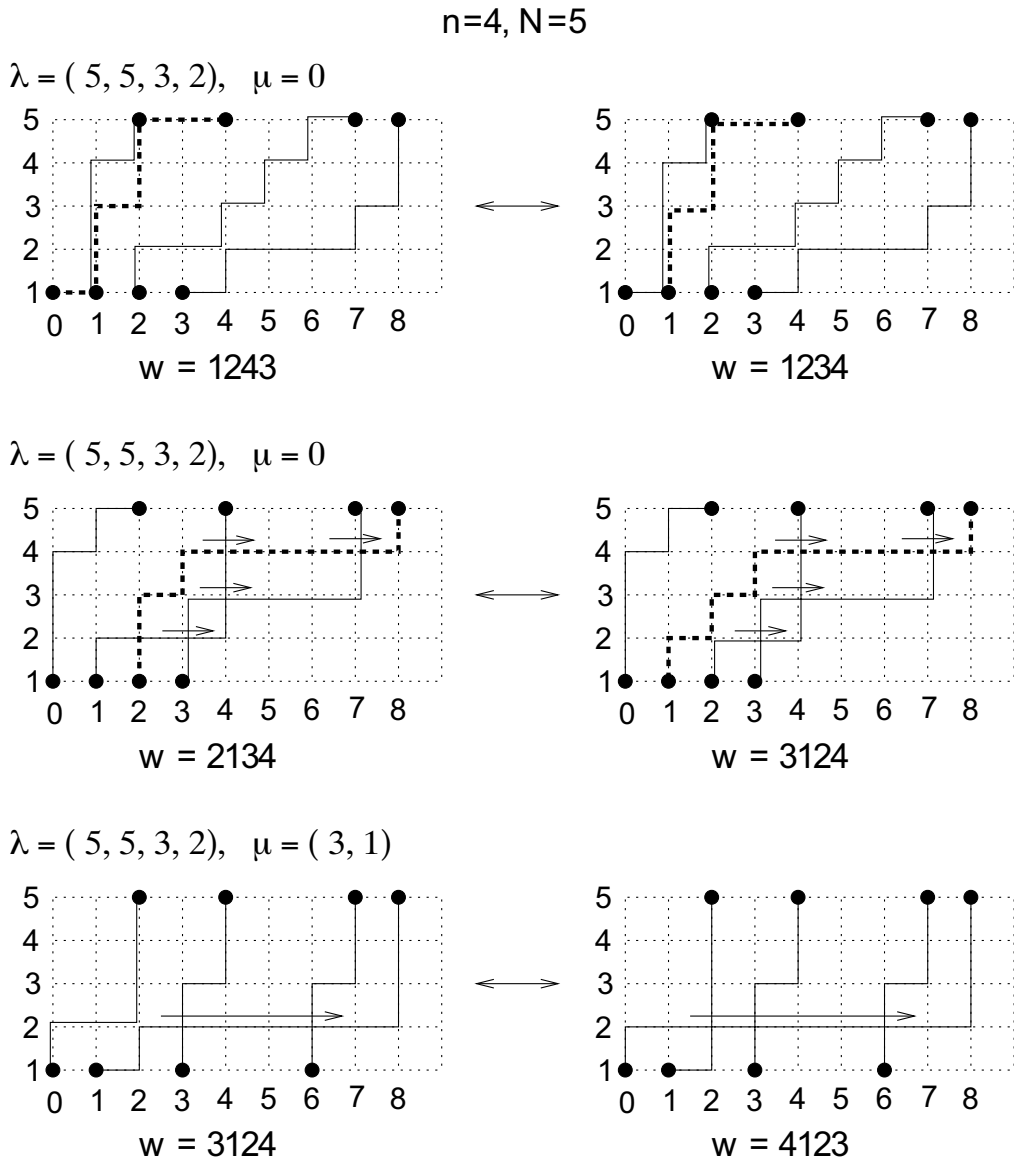
$$\sum_{w \in S_n} \text{sgn}(w) \prod_{i=1}^n h_{\lambda_i - \mu_{w(i)} + w(i) - i}(x_1, \dots, x_N).$$

This is the generating function for the following signed, weighted set. Let  $Z$  be the set of sequences  $(w, p_1, \dots, p_n)$  such that  $w \in S_n$  and  $p_i$  is a path from  $(n - w(i) + \mu_{w(i)}, 1)$  to  $(\lambda_i + n - i, N)$ . The weight of such a sequence is  $\prod_{i=1}^n \text{wt}(p_i)$ , and the sign is  $\text{sgn}(w)$ .

The following involution will cancel all objects  $(w, p_1, \dots, p_n)$  in which two or more paths intersect. Among all lattice points  $(u, v)$  where two paths intersect, choose the one for which  $u$  is minimized; if there are ties, choose the point that minimizes  $v$ . Let  $i < j$  be the two least indices such that  $p_i$  and  $p_j$  pass through  $(u, v)$ . Write  $p_i = qr$  where  $q$  (resp.  $r$ ) is the part of  $p_i$  before (resp. after) the point  $(u, v)$ . Similarly write  $p_j = st$ . Now, pair the given object with the object  $(w', p'_1, \dots, p'_n)$  where  $w' = w \circ (i, j)$ ,  $p'_i = sr$ ,  $p'_j = qt$ , and  $p'_k = p_k$  for all  $k \neq i, j$ . (Thus we have switched the initial segments of the two intersecting paths.) One may check that the new object lies in  $Z$  and has the same weight and opposite sign as the original object. One should also check that applying the map a second time will restore the original object, so we have an involution. Some examples are shown in Figure 11.5. (Note that path  $p_i$  goes from the  $w(i)$ th point from the right on the line  $y = 1$  to the  $i$ th point from the right on the line  $y = N$ .) Let us consider an object  $(w, p_1, \dots, p_n)$  in  $Z$  that is not canceled by the involution. No two paths in this object can intersect. We claim that this forces  $w = \text{id}$ . For otherwise, there would exist  $i < j$  with  $w(i) > w(j)$ . But then  $p_i$  would start to the left of  $p_j$  on the line  $y = 1$  and end to the right of  $p_j$  on the line  $y = N$ ,



**FIGURE 11.4**  
Encoding fillings of a skew shape by sequences of lattice paths.



**FIGURE 11.5**  
Cancellation mechanism for intersecting paths.



which would force  $p_i$  and  $p_j$  to intersect. So  $w = \text{id}$ . Erasing  $w$  maps the fixed points in  $Z$  bijectively to the set  $Y'$ , which in turn maps bijectively to  $\text{SSYT}_N(\lambda/\mu)$ , as shown in the discussion preceding the theorem.  $\square$

**11.61. Theorem: Second Jacobi-Trudi Formula.** Suppose  $\lambda$  is a partition with  $\lambda_1 = n \leq N$ , and  $\mu \subseteq \lambda$ . Then

$$s_{\lambda/\mu}(x_1, \dots, x_N) = \det \|e_{\lambda'_i - \mu'_j + j - i}(x_1, \dots, x_N)\|_{1 \leq i, j \leq n}.$$

*Proof.* For all  $f_{ij} \in \Lambda_N$ , we have  $\omega(\det \|f_{ij}\|) = \det \|\omega(f_{ij})\|$ . This follows from the defining formula for determinants and the fact that  $\omega$  is a ring homomorphism. Now 11.61 follows by applying  $\omega$  to both sides of the first Jacobi-Trudi formula

$$s_{\lambda'/\mu'} = \det \|h_{\lambda'_i - \mu'_j + j - i}\|. \quad \square$$

**11.62. Example.** According to the first Jacobi-Trudi formula,

$$s_{(3,3,1)} = \det \begin{bmatrix} h_3 & h_4 & h_5 \\ h_2 & h_3 & h_4 \\ 0 & 1 & h_1 \end{bmatrix} = h_{(3,3,1)} + h_{(5,2)} - h_{(4,3)} - h_{(4,2,1)}.$$

Note that the main diagonal entries in the formula for  $s_\lambda$  are  $h_{\lambda_1}, h_{\lambda_2}, \dots, h_{\lambda_n}$ , and the subscripts increase by 1 (resp. decrease by 1) as we read to the right (resp. left) along each row. Similarly,

$$s_{(3,3,1)} = \det \begin{bmatrix} e_3 & e_4 & e_5 \\ e_1 & e_2 & e_3 \\ 1 & e_1 & e_2 \end{bmatrix} = e_{(3,2,2)} + e_{(4,3)} + e_{(5,1,1)} - e_{(5,2)} - e_{(3,3,1)} - e_{(4,2,1)}.$$

Here is a typical expansion of a skew Schur polynomial:

$$s_{(5,5,3)/(3,2,0)} = \det \begin{bmatrix} h_2 & h_4 & h_7 \\ h_1 & h_3 & h_6 \\ 0 & 1 & h_3 \end{bmatrix} = h_{(3,3,2)} + h_{(7,1)} - h_{(4,3,1)} - h_{(6,2)}.$$

## 11.15 Inverse Kostka Matrix

In Chapter 10, the Kostka matrix played a prominent role in relating the Schur basis of  $\Lambda_N$  to several other bases. More specifically, we proved the formulas

$$s_\lambda = \sum_{\mu} K_{\lambda, \mu} m_\mu, \quad h_\mu = \sum_{\lambda} K_{\lambda, \mu} s_\lambda, \quad e_\mu = \sum_{\lambda} K_{\lambda, \mu} s_{\lambda'},$$

where all symmetric polynomials have  $N$  variables and all summations extend over  $\text{Par}_N$ . Letting  $\mathbf{K} = \mathbf{K}_N$  be the matrix of Kostka numbers with rows and columns indexed by elements of  $\text{Par}_N$ , these relations can also be written

$$\mathbf{s} = \mathbf{K}\mathbf{m}, \quad \mathbf{h} = \mathbf{K}^t \mathbf{s}, \quad \mathbf{e} = \mathbf{K}^t \omega(\mathbf{s}).$$

We know that the Kostka matrix is invertible (being unitriangular). Let  $K'_{\lambda, \mu}$  be the

entry in row  $\lambda$  and column  $\mu$  of the inverse of the Kostka matrix. Inverting the relations above, we see that

$$m_\lambda = \sum_{\mu} K'_{\lambda,\mu} s_\mu, \quad s_\mu = \sum_{\lambda} K'_{\lambda,\mu} h_\lambda, \quad s_{\mu'} = \sum_{\lambda} K'_{\lambda,\mu} e_\lambda.$$

Observe that the determinant formulas in the previous section, which express Schur polynomials in terms of complete homogeneous symmetric polynomials, give algebraic interpretations for the coefficients  $K'_{\lambda,\mu}$ . Here we wish to derive combinatorial interpretations for these coefficients. To do this, we need the concept of a special rim-hook tableau.

**11.63. Definition: Special Rim-hook Tableaux.** For  $\lambda, \mu \in \text{Par}_N$ , a *special rim-hook tableau of shape  $\mu$  and type  $\lambda$*  is a rim-hook tableau  $S$  of shape  $\mu$  and content  $\alpha$  such that  $\text{sort}(\alpha) = \lambda$  and every nonzero rim-hook in  $S$  contains a cell in the leftmost column of the diagram of  $\mu$ . The *sign* of such a tableau is defined as in 11.48. Let  $\text{SRHT}(\mu, \lambda)$  be the set of special rim-hook tableaux of shape  $\mu$  and type  $\lambda$ .

**11.64. Theorem: Combinatorial Interpretation of Inverse Kostka Matrix.** For all  $\lambda, \mu \in \text{Par}_N$ ,

$$K'_{\lambda,\mu} = \sum_{S \in \text{SRHT}(\mu, \lambda)} \text{sgn}(S).$$

*Proof.* We intend to give a combinatorial proof of the identity

$$a_{\delta(N)}(x_1, \dots, x_N) m_\lambda(x_1, \dots, x_N) = \sum_{\mu \in \text{Par}_N} \sum_{S \in \text{SRHT}(\mu, \lambda)} \text{sgn}(S) a_{\mu + \delta(N)}(x_1, \dots, x_N).$$

Once this is done, the theorem will follow by dividing both sides by  $a_{\delta(N)}$  and comparing the resulting identity to the known expansion  $m_\lambda = \sum_{\mu} K'_{\lambda,\mu} s_\mu$ .

To prove the identity, we study a combinatorial interpretation of the product  $a_{\delta(N)} m_\lambda$  involving abaci. The polynomial  $a_{\delta(N)}$  represents a justified abacus containing  $N$  beads labeled  $w(N), \dots, w(1)$  in positions  $0, \dots, N-1$  (respectively). Given such an abacus, we can view  $m_\lambda(x_1, \dots, x_N)$  as the sum of all distinct monomials  $\prod_{i=1}^N x_{w(i)}^{e(i)}$  such that the exponent sequence  $(e(1), \dots, e(N))$  is a rearrangement of  $(\lambda_1, \dots, \lambda_N)$ . (Here and below, we view elements of  $\text{Par}_N$  as partitions with *exactly*  $N$  parts, some of which may be zero.) The multiplication of  $a_{\delta(N)}$  by one of these monomials can be implemented on the abacus as follows. Imagine moving the  $N$  justified beads from their current runner to a new, initially empty runner, by moving each bead  $w(i)$  from position  $N-i$  on the old runner to position  $N-i+e(i)$  on the new runner. Call such a transformation of the justified abacus a  $\lambda$ -move. A given  $\lambda$ -move either causes a bead collision on the new runner, or else produces a new abacus, which is enumerated by a monomial in  $a_{\mu + \delta(N)}(x_1, \dots, x_N)$  for some  $\mu \in \text{Par}_N$ .

Consider the situation where a bead collision occurs. Choose  $i$  minimal such that bead  $w(i)$  collides with some other bead on the new runner, and then choose  $j$  minimal such that bead  $w(i)$  collides with  $w(j)$ . Create a new object counted by  $a_{\delta(N)} m_\lambda$  by switching beads  $w(i)$  and  $w(j)$  on the old abacus, and switching  $e(i)$  and  $e(j)$  in the exponent vector. This defines a sign-reversing, weight-preserving involution that cancels all objects in which bead collisions occur.

To complete the proof, we must find a sign-preserving, weight-preserving bijection  $\phi$  from the set  $X$  of uncanceled objects counted by  $a_{\delta(N)} m_\lambda$  to the signed weighted set

$$\bigcup_{\mu \in \text{Par}_N} \text{SRHT}(\mu, \lambda) \times \text{LAbc}(\mu + \delta(N)).$$

$$S = \begin{array}{|c|c|c|c|c|c|} \hline 1 & 2 & 2 & 2 & 2 & 2 \\ \hline 2 & 2 & 4 & 4 & 4 & \\ \hline 4 & 4 & 4 & 5 & & \\ \hline 4 & 5 & 5 & 5 & & \\ \hline 5 & 5 & & & & \\ \hline \end{array}$$
**FIGURE 11.6**

A special rim-hook tableau.

For this purpose, let us fix  $\mu \in \text{Par}_N$  and consider the ways in which a justified abacus with  $N$  beads can be transformed into an abacus in  $\text{LAbc}(\mu + \delta(N))$  by means of a  $\lambda$ -move. Let us temporarily ignore bead labels and signs, concentrating at first only on the positions of the  $N$  beads. The positions of the  $N$  beads on the old runner are the entries in the sequence  $\delta(N) = (N - 1, N - 2, \dots, 2, 1, 0)$ . A  $\lambda$ -move adds some rearrangement of the sequence  $\lambda = (\lambda_1, \dots, \lambda_N)$  to the sequence  $\delta(N)$ . We will obtain an abacus in  $\text{LAbc}(\mu + \delta(N))$  iff the sum of these sequences is some rearrangement of the sequence  $\mu + \delta(N) = (\mu_1 + N - 1, \dots, \mu_N + N - N)$ .

We now show that the rearrangements of  $\lambda$  that produce abaci in  $\text{LAbc}(\mu + \delta(N))$  can be encoded by special rim-hook tableaux of shape  $\mu$  and type  $\lambda$ . The proof will use induction on  $N$ . Let us first illustrate the idea of the proof by considering an example. Take  $N = 5$ ,  $\mu = (7, 5, 4, 4, 2)$ , and  $\lambda = (8, 7, 6, 1, 0)$ . We seek rearrangements of the vector  $(8, 7, 6, 1, 0)$  which, when added to the vector  $(4, 3, 2, 1, 0)$ , produce a rearrangement of  $\mu + \delta(N) = (11, 8, 6, 5, 2)$ . In this example, the only solution turns out to be  $(1, 8, 0, 7, 6) + (4, 3, 2, 1, 0) = (5, 11, 2, 8, 6)$ . We can visualize this solution using the special rim-hook tableau in Figure 11.6, in which the rim-hooks (from top to bottom) have lengths  $(1, 8, 0, 7, 6)$ . If we start with a labeled justified abacus  $54321000\dots$  and perform a  $\lambda$ -move using the rearrangement  $(1, 8, 0, 7, 6)$ , we obtain the abacus  $003001504002000\dots \in \text{LAbc}(11, 8, 6, 5, 2)$ . The sign of this abacus, namely  $\text{sgn}(24513) = -1$ , differs from the sign of the original abacus, namely  $\text{sgn}(12345) = +1$ , by a factor of  $(-1)^5 = \text{sgn}(S)$ . A similar remark holds if the original abacus had involved some other permutation of the five labels.

With this example in mind, we return to the general proof. We are seeking permutations  $j_1 \dots j_N$  and  $k_1 \dots k_N$  satisfying the system of equations

$$\begin{aligned} 0 + \lambda_{j_N} &= \mu_{k_N} + N - k_N \\ 1 + \lambda_{j_{N-1}} &= \mu_{k_{N-1}} + N - k_{N-1} \\ &\dots &\dots \\ N - 1 + \lambda_{j_1} &= \mu_{k_1} + N - k_1 \end{aligned} \tag{11.5}$$

In particular, to satisfy the first equation, we need an index  $j = j_N$  and an index  $k = k_N$  such that  $\lambda_j = \mu_k + N - k$ . If such an index exists, we encode it by drawing the unique border ribbon of length  $\lambda_j$  starting in the leftmost cell of row  $N$  of  $\mu$ . By choice of  $j$  and  $k$ , this border ribbon must end in the rightmost cell of row  $k$  of  $\mu$ . In terms of the abaci, the  $\lambda$ -move encoded by  $j_1 \dots j_N$  moves the bead in position 0 on the old runner (the  $N$ th bead from the right) to position  $\mu_k + N - k$  on the new runner (which will become the  $k$ th bead from the right). Thus this bead “moves past”  $N - k$  other beads during the  $\lambda$ -move, which causes a sign change of  $(-1)^{N-k}$  for any choice of labels. But  $N - k$  is precisely the spin of the border ribbon we just drew.

To finish solving system (11.5), let  $\lambda^*$  be the partition obtained by dropping one part  $\lambda_j$  from  $\lambda$ , and let  $\mu^*$  be the partition in  $\text{Par}_{N-1}$  obtained by erasing the cells of  $\mu$  occupied by

the ribbon that starts in row  $N$ . Suppose we ignore the first equation in the system (11.5) and subtract 1 from both sides of the remaining  $N - 1$  equations. One may check that the resulting system of  $N - 1$  equations is precisely the system we must solve to change a justified abacus to an abacus in  $\text{LAbc}(\mu^* + \delta(N - 1))$  by means of a  $\lambda^*$ -move. (For instance, in the example considered earlier, after we move a bead from position 0 to position 6 [accounting for the lowest rim-hook in the displayed tableau], we have  $\lambda^* = (8, 7, 1, 0)$  and  $\mu^* = (7, 5, 3, 1)$ . Having moved one bead, we are left with the task of moving beads from positions  $(4, 3, 2, 1) = (1, 1, 1, 1) + \delta(4)$  to positions  $(11, 8, 5, 2) = (1, 1, 1, 1) + \mu^* + \delta(4)$  using the moves in  $\lambda^* = (8, 7, 1, 0)$ .) By induction on  $N$ , the solutions of the reduced system are encoded by special rim-hook tableaux  $S^*$  of shape  $\mu^*$  and type  $\lambda^*$ ; and furthermore, the net sign change going from the old abacus to the new abacus (disregarding the bead originally in position 0) is  $\text{sgn}(S^*)$ . It follows that all solutions of the original system are encoded by special rim-hook tableaux  $S$  of shape  $\mu$  and type  $\lambda$ ; and furthermore, the net sign change going from the old abacus to the new abacus (taking all beads into account) is  $\text{sgn}(S)$ .

The preceding discussion contains an implicit recursive definition of the desired bijection  $\phi$ . More explicitly, suppose  $z = (w(N) \cdots w(1)000 \cdots, e(N) \cdots e(1)) \in X$  is an uncanceled object counted by  $a_{\delta(N)} m_\lambda$ . Then  $\phi(z) = (S, v)$  where  $v \in \text{LAbc}(\mu + \delta(N))$  is obtained from the first component of  $z$  by moving bead  $w(i)$  right  $e(i)$  positions for all  $i$ , and  $S$  is the unique special rim-hook tableau (of shape  $\mu$  determined by  $v$ ) that has a rim-hook of length  $e(i)$  starting in the leftmost cell of row  $i$  of the diagram. The preceding arguments show that  $\phi$  preserves signs and weights. To compute  $\phi^{-1}(S, v)$ , it suffices to note that the sequence  $(e(1), \dots, e(N))$  is the content of the rim-hook tableau  $S$ . Knowledge of this sequence allows us to reverse the  $\lambda$ -move and recover  $w(N) \cdots w(1)$ . Thus,  $\phi$  is a bijection.  $\square$

**11.65. Remark.** An alternate approach to the theorem is to *define*

$$K'_{\lambda, \mu} = \sum_{S \in \text{SRHT}(\mu, \lambda)} \text{sgn}(S)$$

and then give a combinatorial proof of the matrix identity  $\mathbf{K}\mathbf{K}' = \mathbf{I}$  (see 11.127). Since  $\mathbf{K}$  is known to be invertible, it follows that  $\mathbf{K}'$  must be the (two-sided) matrix inverse of  $\mathbf{K}$ .

## 11.16 Schur Expansion of Skew Schur Polynomials

We now consider the expansion of skew Schur polynomials as linear combinations of ordinary Schur polynomials. Since the ordinary Schur polynomials are a basis of  $\Lambda_N$  and the skew Schur polynomials are in this vector space, we know there exist unique scalars  $c_{\nu, \mu}^\lambda \in \mathbb{Q}$  such that

$$s_{\lambda/\nu}(x_1, \dots, x_N) = \sum_{\mu} c_{\nu, \mu}^\lambda s_{\mu}(x_1, \dots, x_N), \quad (11.6)$$

where it suffices to sum over partitions  $\mu$  of size  $|\lambda/\nu|$ . The scalars  $c_{\nu, \mu}^\lambda$  are called *Littlewood-Richardson coefficients*. The following result shows that these coefficients are all nonnegative integers. Recall that, for a semistandard tableau  $T$  of any shape, the *word* of  $T$  is obtained by concatenating the rows of  $T$  from bottom to top.

**11.66. Theorem: Littlewood-Richardson Rule for Skew Schur Polynomials.** For all partitions  $\lambda, \mu, \nu$ ,  $c_{\nu, \mu}^\lambda$  is the number of semistandard tableaux  $T$  of shape  $\lambda/\nu$  and content  $\mu$  such that every suffix of the word of  $T$  has partition content. In other words,

writing  $w(T) = w_1 w_2 \cdots w_n$ , we require that for all  $k \leq n$  and all  $i \geq 1$ , the number of  $i$ 's in the suffix  $w_k w_{k+1} \cdots w_n$  equals or exceeds the number of  $i + 1$ 's in this suffix.

*Proof.* Multiplying both sides of (11.6) by  $a_{\delta(N)}$ , it suffices to prove the identity

$$a_{\delta(N)}(x_1, \dots, x_N) s_{\lambda/\nu}(x_1, \dots, x_N) = \sum_{\mu} c_{\nu, \mu}^{\lambda} a_{\mu + \delta(N)}(x_1, \dots, x_N).$$

The idea is to generalize the proof of the special case  $\nu = (0)$  which we gave in §11.12. Model the left side of the desired identity by the set  $X$  of pairs  $(v, T)$ , where  $v$  is a justified  $N$ -bead labeled abacus and  $T$  is a semistandard tableau of shape  $\lambda/\nu$  over the alphabet  $\{1, 2, \dots, N\}$  ordered by  $<_v$ . Since skew Schur polynomials are symmetric, the generating function for the signed, weighted set  $X$  is  $a_{\delta(N)} s_{\lambda/\nu}$ .

We now define an involution  $I : X \rightarrow X$ . Given  $(v, T) \in X$ ,  $T$  determines a sequence of bead motions on  $v$  by reading  $w(T)$  from right to left and moving bead  $k$  one step to the right each time the symbol  $k$  is seen. If these bead motions cause a collision, define  $I(v, T) = (v', T')$  by the following steps. Suppose the first collision occurs when bead  $i$  bumps into bead  $j$  (where  $i >_v j$  are adjacent beads in  $v$ ). Let  $v'$  be  $v$  with beads  $i$  and  $j$  switched, so  $\text{sgn}(v') = -\text{sgn}(v)$  and  $\text{wt}(v')x_j = \text{wt}(v)x_i$ .

Next, we calculate  $T'$  from  $T$  as follows. Starting with the word of  $T$ , replace each  $i$  by a left parenthesis, each  $j$  by a right parenthesis, and ignore all other symbols. Match left and right parentheses in the resulting string of parentheses, and ignore these matched pairs of parentheses hereafter. The remaining unmatched parentheses must consist of a string of  $a \geq 0$  right parentheses followed by a string of  $b \geq 0$  left parentheses, since if a left parenthesis appeared somewhere to the left of a right parenthesis we could find another matched pair of parentheses.

Note that  $b > 0$ , since otherwise bead  $i$  would never bump into bead  $j$ . Indeed, the first bead collision occurs when we reach the rightmost unmatched left parenthesis (occurrence of  $i$ ) in the word of  $T$ . Now, change the subword of unmatched parentheses from “ $)^a(b^b$ ” to “ $)^{b-1}(a+1$ ”, and then convert all left parentheses to  $j$ 's and all right parentheses to  $i$ 's. One may verify that the new word is the word of a tableau  $T' \in \text{SSYT}_N(\lambda/\nu)$ , relative to the ordering  $<_{v'}$ , using the facts that  $i$  and  $j$  are adjacent relative to the orderings  $<_v$  and  $<_{v'}$ , and the status of a given parenthesis symbol in  $T'$  (matched or unmatched) is the same as its status in  $T$ . See the example following the proof for more discussion of this point.

Because  $T'$  has one less  $i$  than  $T$  and  $T'$  has one more  $j$  than  $T$ , we have  $\text{wt}(T')x_i = \text{wt}(T)x_j$ . Since we also had  $\text{wt}(v')x_j = \text{wt}(v)x_i$ , we see that  $\text{wt}(v', T') = \text{wt}(v, T)$ . Thus  $I$  is sign-reversing and weight-preserving. Finally, to check that  $I$  is an involution, consider what happens when we use  $T'$  to move the beads on  $v'$ . Bead  $j$  on  $v'$  moves the same way as bead  $i$  did on  $v$  (and vice versa) until we reach the rightmost unmatched parenthesis (relative to  $i$  and  $j$ ) in  $w(T')$ . When this symbol is reached, bead  $j$  bumps into bead  $i$  on  $v'$ , just as bead  $i$  bumped into bead  $j$  on  $v$ . To compute  $I(v', T')$ , we will therefore apply the parenthesis modification rule to the  $i$ 's and  $j$ 's appearing in  $w(T')$ . This rule will change the unmatched parentheses from “ $)^{b-1}(a+1$ ” back to “ $)^a(b^b$ ”, which shows that  $I(v', T') = (v, T)$ . So  $I$  is an involution.

All that remains is to analyze the fixed points of  $I$ , which are (by definition) the pairs  $(v, T)$  for which no bead collision occurs. Recall that we are starting with a justified abacus  $v$ , scanning the symbols in  $w(T) = w_1 \cdots w_n$  from right to left, and moving the corresponding beads on  $v$ . Suppose all suffixes of  $T$  have partition content relative to the ordering  $<_v$  (which means the rightmost bead label occurs at least as often in each suffix as the next bead label, and so on). We see from the description of the bead motion that no collision will occur. Conversely, if the condition is first violated by some suffix  $w_k w_{k+1} \cdots w_n$ , then a collision will occur at this point in the scan. Thus the fixed points of  $I$  are the pairs  $(v, T)$

such that each suffix of  $T$  has partition content relative to  $<_v$ . We map each such fixed point to the abacus  $v^*$  obtained from  $v$  by performing the bead motions specified by  $T$ . The abacus  $v^*$  lies in the set  $\text{LAbc}(\mu + \delta(N))$ , where  $\mu$  is the content of  $T$  (calculating content relative to  $<_v$ , so  $\mu_1$  is the number of times the rightmost bead moves, etc.).

We can obtain all the fixed points of  $I$  from fixed points of the form  $(v^0, T)$ , where  $v^0 = (N, N-1, \dots, 1, 0, 0, \dots)$ ,  $<_{v^0}$  is the usual ordering on integers, and  $T$  is a semistandard tableau satisfying the conditions in the theorem statement. We need only permute the bead labels in  $v^0$  by any  $w \in S_N$ , and permute the entries of  $T$  in the same way. The object  $(v^0, T)$  thereby generates  $N!$  fixed points which together contribute one copy of  $a_{\mu+\delta(N)}(x_1, \dots, x_N)$  to the generating function for the fixed points of  $I$ . The total number of times this term appears in the generating function is the total number of semistandard tableaux  $T$  of content  $\mu$  satisfying the conditions in the theorem. Since the generating function for  $X$  must equal the generating function for the fixed point set of  $I$ , the proof is complete.  $\square$

**11.67. Example.** To illustrate the parenthesis construction, we compute  $I(5432100 \dots, T)$ , where

$$T = \begin{array}{cccccccccccccccc} & & & & & & & & & & & & & & & & 1 & 1 & 1 \\ & & & & & & & & & & & & & & & 1 & 2 & 2 & 2 & 2 \\ & & & & & & & & & & & & & & 1 & 1 & 1 & 1 & 2 \\ & & & & & & & & & & & & 1 & 1 & 1 & 2 & 2 & 2 \\ & & & & & & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ & & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ & 1 & 2 & 2 \end{array}.$$

The word of  $T$  is 12211122221111212222111. The suffix 2222111 of  $w(T)$  does not have partition content, so this object will cancel with some object  $(5431200 \dots, T')$ . To find  $T'$ , first convert 1's to right parentheses and 2's to left parentheses in  $w(T)$ :

12211122221111212222111  
 $)((()))((((( )))((((( )))$

Now we balance parentheses and mark the remaining unmatched symbols:

$)((()))((((( )))((((( )))$   
 $* \quad * \quad \quad *$

The substring of unmatched parentheses is “ $)()$ .” Observe that the rightmost symbol in this substring is a left parenthesis corresponding to the first 2 in the offending suffix 2222111, and this 2 is the symbol in  $w(T)$  causing the first bead collision. As directed by the proof, we convert the unmatched parenthesis string to “ $((()$ ” and then replace left parentheses by 1's and right parentheses by 2's:

$* \quad * \quad \quad *$   
 $((()))((((( )))((((( )))$   
 11122111112222121111222

This new word is  $w(T')$ , so finally

$$T' = \begin{array}{cccccccccccccccc} & & & & & & & & & & & & & & & & 2 & 2 & 2 \\ & & & & & & & & & & & & & & & 2 & 1 & 1 & 1 & 1 \\ & & & & & & & & & & & & & & 2 & 2 & 2 & 2 & 1 \\ & & & & & & & & & & & & 2 & 2 & 1 & 1 & 1 & 1 & 1 \\ & & & & & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 & 1 \end{array}.$$

Observe that  $T'$  is a semistandard tableau relative to the ordering  $5 > 4 > 3 > 1 > 2$ . In particular, columns of  $T'$  strictly increase because whenever 1 appears above 2 in  $T$ , these occurrences of 1 and 2 become matched parentheses. Rearranging the unmatched parentheses does not affect these symbols, so in the end we will get a 2 above a 1 in  $T'$ . Also,

rows of  $T'$  weakly increase since a strict decrease in some row would be encoded as a matched parenthesis pair in  $w(T')$ , which would have also been matched in  $w(T)$ , implying that  $T$  had a strict decrease in some row. But  $T$  is a semistandard tableau so this cannot happen. Finally, note that the shortest suffix of  $T'$  that does not have partition content (relative to the new ordering) is 1111222, where the leftmost 1 corresponds to the rightmost unbalanced parenthesis in  $w(T')$ . Consequently,  $I(5432100 \cdots, T') = (5432100 \cdots, T)$ . Observe that these two objects have opposite sign, but both have weight  $x_1^{16} x_2^{14} x_3^2 x_4^1 x_5^0$ .

**11.68. Example.** Let us compute  $I(v^0, T)$ , where

$$v_0 = 5432100 \cdots, \quad T = \begin{array}{cccc} & & & 1 & 1 & 1 \\ & & & 2 & 3 & 3 \\ 1 & 1 & 2 & 4 & 4 & 5 \\ 2 & 2 & 3 & 5 & & \\ 3 & 5 & & & & \end{array}$$

Moving beads on  $v^0$  according to the word  $w(T) = 352235112445233111$ , bead 3 bumps into bead 2 when we have scanned the suffix 3111 (which is the shortest suffix without partition content). We therefore modify the 2's and 3's in the word as follows:

352235112445233111  
3 223    2    233  
( ) (    )    ) ((  
     \*       \*\*\*  
( ) (    )    ((  
2 332    3    222  
253325113445222111

Therefore  $I(v^0, T) = (v', T')$ , where

$$v' = 54231, \quad T' = \begin{array}{cccc} & & & 1 & 1 & 1 \\ & & & 2 & 2 & 2 \\ 1 & 1 & 3 & 4 & 4 & 5 \\ 3 & 3 & 2 & 5 & & \\ 2 & 5 & & & & \end{array}$$

Observe that  $\text{wt}(v^0, T) = \text{wt}(v', T') = x_1^9 x_2^7 x_3^6 x_4^3 x_5^3$ ,  $\text{sgn}(v', T') = -\text{sgn}(v^0, T)$ , and  $I(v', T') = (v^0, T)$ .

**11.69. Example.** Let us compute  $c_{\nu, \mu}^\lambda$  when  $\lambda = (5, 4, 4, 1)$ ,  $\nu = (3, 1)$ , and  $\mu = (4, 4, 2)$ . We draw the semistandard tableaux of shape  $\lambda/\nu$  whose words have the required suffix property. The following two tableaux are the only ones, so  $c_{\nu, \mu}^\lambda = 2$ :

$$T_1 = \begin{array}{cccc} & & & 1 & 1 \\ & & 1 & 1 & 2 \\ 2 & 2 & 2 & 3 & \\ 3 & & & & \end{array}, \quad T_2 = \begin{array}{cccc} & & & 1 & 1 \\ & & 1 & 2 & 2 \\ 1 & 2 & 3 & 3 & \\ 2 & & & & \end{array}$$

Let us see how these tableaux correspond to fixed points of  $I$  when  $N = 5$ . The first tableau changes the standard abacus  $v^0 = (5432100 \cdots)$  to the abacus  $(54003002100 \cdots)$  counted by  $\text{LAbc}(\mu + \delta(5))$  by moving bead 1 twice, then bead 2 once, then bead 1 twice, and so on. Permuting the labels gives the other 119 signed objects that make up one copy of  $a_{\mu+\delta(5)}(x_1, \dots, x_5)$ ; for instance,

$$\left( 3425100 \cdots, \begin{array}{cccc} & & & 1 & 1 \\ & & 1 & 1 & 5 \\ 5 & 5 & 5 & 2 & \\ 2 & & & & \end{array} \right) \mapsto (34002005100 \cdots).$$

On the other hand, the second tableau changes the standard abacus  $(5432100\cdots)$  to the abacus  $(54003002100\cdots)$  via a different sequence of collision-free bead moves: move bead 1 twice, then bead 2 twice, then bead 1 once, and so on. This pair and its permutations produce another copy of the generating function  $a_{\mu+\delta(5)}(x_1, \dots, x_5)$ . Dividing by  $a_{\delta(5)}$ , we conclude that

$$s_{\lambda/\nu} = 2s_\mu + \cdots.$$

Now let us compute  $c_{\mu,\nu}^\lambda$ . The required skew tableaux, which have shape  $(5, 4, 4, 1)/(4, 4, 2)$  and content  $(3, 1)$ , are:

$$\begin{array}{c} \boxed{1} \\ \boxed{2} \end{array} \begin{array}{c} \boxed{1} \boxed{1} \\ \boxed{1} \boxed{1} \end{array}, \quad \begin{array}{c} \boxed{1} \\ \boxed{1} \end{array} \begin{array}{c} \boxed{1} \boxed{2} \\ \boxed{1} \boxed{2} \end{array}.$$

So  $c_{\mu,\nu}^\lambda = 2$ . This illustrates the general symmetry property  $c_{\nu,\mu}^\lambda = c_{\mu,\nu}^\lambda$ , which is true but not immediately evident from our combinatorial description of these coefficients. We will prove this property later when we discuss products of Schur polynomials.

**11.70. Example.** For  $N \geq 7$ , let us find the Schur expansion of the skew Schur polynomial  $s_{(3,3,2,2)/(2,1)}$  in  $N$  variables. This expansion is found by enumerating all semistandard skew tableaux of shape  $(3, 3, 2, 2)/(2, 1)$  satisfying the required suffix property. Each such tableau of content  $\mu$  contributes one term  $s_\mu$  to the expansion. The relevant tableaux are shown here:

$$\begin{array}{c} \boxed{1} \\ \boxed{1} \boxed{2} \\ \boxed{1} \boxed{2} \boxed{3} \end{array} \quad \begin{array}{c} \boxed{1} \\ \boxed{1} \boxed{2} \\ \boxed{3} \boxed{3} \end{array} \quad \begin{array}{c} \boxed{1} \\ \boxed{1} \boxed{2} \\ \boxed{1} \boxed{3} \boxed{4} \end{array} \quad \begin{array}{c} \boxed{1} \\ \boxed{1} \boxed{2} \\ \boxed{2} \boxed{3} \boxed{4} \end{array}$$

We conclude that

$$s_{(3,3,2,2)/(2,1)} = 1s_{(3,3,1)} + 1s_{(3,2,2)} + 1s_{(3,2,1,1)} + 1s_{(2,2,2,1)}.$$

## 11.17 Products of Schur Polynomials

Given partitions  $\mu \in \text{Par}_N(m)$  and  $\nu \in \text{Par}_N(n)$ , the product  $s_\mu(x_1, \dots, x_N)s_\nu(x_1, \dots, x_N)$  is a symmetric polynomial, so it can be expressed uniquely in terms of Schur polynomials  $s_\lambda(x_1, \dots, x_N)$  indexed by  $\lambda \in \text{Par}_N(m+n)$ :

$$s_\mu s_\nu = \sum_{\lambda} a(\lambda, \mu, \nu) s_\lambda \quad (a(\lambda, \mu, \nu) \in \mathbb{Q}). \quad (11.7)$$

Now, 11.57 shows that the coefficients here are precisely the Littlewood-Richardson numbers:

$$a(\lambda, \mu, \nu) = \langle s_\mu s_\nu, s_\lambda \rangle = \langle s_\mu, s_{\lambda/\nu} \rangle = c_{\nu,\mu}^\lambda.$$

Since  $s_\mu s_\nu = s_\nu s_\mu$  (because multiplication of polynomials is commutative), we deduce the symmetry

$$c_{\nu,\mu}^\lambda = c_{\mu,\nu}^\lambda.$$

We now derive another combinatorial expression for these integers by viewing the product  $s_\mu s_\nu$  as a skew Schur polynomial. We claim that  $s_\mu s_\nu = s_{\alpha/\beta}$ , where

$$\alpha = (\mu_1 + \nu_1, \mu_1 + \nu_2, \dots, \mu_1 + \nu_N, \mu_1, \dots, \mu_N), \quad \beta = (\mu_1^N).$$

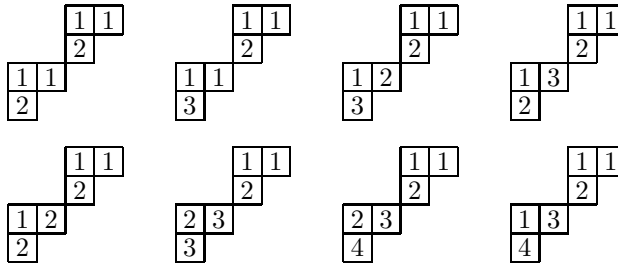


This follows since the skew shape  $\alpha/\beta$  consists of two disconnected pieces, one of shape  $\nu$  and one of shape  $\mu$ . A semistandard skew tableau of this shape can be formed by choosing a semistandard tableau of shape  $\nu$  and independently choosing a semistandard tableau of shape  $\mu$ ; thus the result follows from the product rule for weighted sets. We conclude that

$$c_{\mu,\nu}^{\lambda} = c_{\beta,\lambda}^{\alpha} \quad (\text{with } \alpha, \beta \text{ as above}).$$

This formula is illustrated in the next example.

**11.71. Example.** Let us compute the Schur expansion of  $s_{(2,1)}s_{(2,1)}$  using the observation  $s_{(2,1)}s_{(2,1)} = s_{(4,3,2,1)/(2,2)}$ . The following skew tableaux have words such that all suffixes have partition content:



Looking at contents, we conclude that

$$s_{(2,1)}s_{(2,1)} = s_{(4,2)} + s_{(4,1,1)} + 2s_{(3,2,1)} + s_{(3,3)} + s_{(2,2,2)} + s_{(2,2,1,1)} + s_{(3,1,1,1)}.$$

Observe that the upper-right portion of the skew tableau could only be filled in one way. So we could ignore this part of the tableau and just consider suitable fillings of the lower shape. Generalizing this remark leads to the following prescription for the Littlewood-Richardson coefficients.

**11.72. Theorem: Alternate Formula for Littlewood-Richardson Coefficients.** For all partitions  $\lambda, \mu, \nu \in \text{Par}_N$ , the coefficient  $c_{\mu,\nu}^{\lambda} = c_{\beta,\lambda}^{\alpha}$  is the number of semistandard tableaux  $T$  of shape  $\mu$  and content  $\lambda - \nu = (\lambda_i - \nu_i : 1 \leq i \leq N)$  such that  $w(T) = w_1 \cdots w_n$  satisfies the following condition: for all  $k \leq n$ , the exponent vector of the monomial  $\prod_{j=1}^N x_j^{\nu_j} \prod_{i=k}^n x_{w_i}$  is a partition. (This condition means that for all  $k \leq n$  and  $j < N$ , if there are  $a$  copies of  $j$  and  $b$  copies of  $j+1$  in the suffix  $w_k w_{k+1} \cdots w_n$ , then  $\nu_j + a \geq \nu_{j+1} + b$ .)

*Proof.* We already know that  $c_{\mu,\nu}^{\lambda} = c_{\beta,\lambda}^{\alpha}$  where the skew shape  $\alpha/\beta$  consists of an upper part of shape  $\nu$  and a lower part of shape  $\mu$ . We also know that  $c_{\beta,\lambda}^{\alpha}$  is the number of skew tableaux  $U$  of shape  $\alpha/\beta$  and content  $\lambda$  such that every suffix of  $w(U)$  has partition content. Consider the last  $|\nu|$  symbols in  $w(U)$ . The last letter is the label in the rightmost cell of the first row of the skew shape  $\alpha/\beta$ . The partition content condition forces this letter to be 1, and then all letters in the first row of the skew tableau  $U$  must be 1. The letter at the end of the second row must be strictly greater than 1, so it is 2 (by the partition content condition), and then every entry in the second row must be 2. Proceeding in this way, we see that for  $k \leq N$ , every entry in row  $k$  of  $U$  must equal  $k$ . Equivalently, the last  $|\nu|$  symbols of  $w(U)$  must be  $N^{\nu_N} \cdots 2^{\nu_2} 1^{\nu_1}$ . Call this suffix  $z$ .

Now we must fill the lower part of the shape  $\alpha/\beta$  by choosing a semistandard tableau  $T$  of shape  $\mu$ . Because the upper part of  $U$  has content  $\nu$ , the content of the entire skew tableau  $U$  will be  $\lambda$  iff the content of the lower part  $T$  is  $\lambda - \nu$ . The other condition imposed on  $T$  is that, for every suffix  $y$  of  $w(T)$ , the suffix  $yz$  of  $w(U)$  has partition content. Given the formula for  $z$  above, this condition is equivalent to the condition on  $w(T)$  in the theorem statement.  $\square$

**TABLE 11.1**

Formulas for manipulating antisymmetric and symmetric polynomials.

Pieri rules:	$a_{\lambda+\delta(N)}p_k = \sum_{\beta: \beta/\lambda \text{ is a } k\text{-ribbon } R} \text{sgn}(R)a_{\beta+\delta(N)};$ $a_{\lambda+\delta(N)}e_k = \sum_{\beta: \beta/\lambda \text{ is a vertical } k\text{-strip}} a_{\beta+\delta(N)};$ $a_{\lambda+\delta(N)}h_k = \sum_{\beta: \beta/\lambda \text{ is a horizontal } k\text{-strip}} a_{\beta+\delta(N)};$ $s_{\lambda}p_k = \sum_{\beta: \beta/\lambda \text{ is a } k\text{-ribbon } R} \text{sgn}(R)s_{\beta};$ $s_{\mu}p_{\alpha} = \sum_{\lambda} \chi_{\alpha}^{\lambda/\mu} s_{\lambda}.$
Determinant formula for $s_{\lambda}$ :	$s_{\lambda} = \frac{a_{\lambda+\delta(N)}}{a_{\delta(N)}} = \frac{\det \ x_j^{\lambda_i+N-i}\ _{1 \leq i, j \leq N}}{\det \ x_j^{N-i}\ _{1 \leq i, j \leq N}}.$
Schur expansion of power-sums:	$p_{\alpha} = \sum_{\lambda} \chi_{\alpha}^{\lambda} s_{\lambda}.$
Power-sum expansion of Schur polys.:	$s_{\lambda} = \sum_{\mu} \frac{\chi_{\mu}^{\lambda}}{z_{\mu}} p_{\mu}.$
Formulas for skew Schur polynomials:	$s_{\lambda/\mu} = \sum_{\nu} \frac{\chi_{\nu}^{\lambda/\mu}}{z_{\nu}} p_{\nu};$ $\langle s_{\lambda/\mu}, f \rangle = \langle s_{\lambda}, s_{\mu}f \rangle \text{ for } f \in \Lambda_N;$ $\omega(s_{\lambda/\mu}) = s_{\lambda'/\mu'};$ $s_{\lambda/\mu} = \det \ h_{\lambda_i - \mu_j + j - i}\ _{1 \leq i, j \leq \ell(\lambda)};$ $s_{\lambda/\mu} = \det \ e_{\lambda'_i - \mu'_j + j - i}\ _{1 \leq i, j \leq \lambda_1};$ $s_{\lambda/\mu} = \sum_{\nu} c_{\mu, \nu}^{\lambda} s_{\nu};$ $s_{\mu} s_{\nu} = \sum_{\lambda} c_{\lambda, \mu, \nu}^{\lambda} s_{\lambda}.$
Inverse Kostka formulas:	$(K'_{\lambda, \mu} = \sum_{S \in \text{SRHT}(\mu, \lambda)} \text{sgn}(S))$ $m_{\lambda} = \sum_{\mu} K'_{\lambda, \mu} s_{\mu};$ $s_{\mu} = \sum_{\lambda} K'_{\lambda, \mu} h_{\lambda};$ $s_{\mu'} = \sum_{\lambda} K'_{\lambda, \mu} e_{\lambda}.$

## Summary

Table 11.1 contains formulas derived in this chapter for computing with antisymmetric and symmetric polynomials.

- *Unlabeled Abaci.* An abacus is a function  $w : \mathbb{Z} \rightarrow \{0, 1\}$  with  $w(i) = 1$  for all small enough  $i$  and  $w(j) = 0$  for all large enough  $j$ . Justification of abaci gives a bijection to pairs  $(m, \lambda) \in \mathbb{Z} \times \text{Par}$ . The inverse bijection can be computed by traversing the frontier of  $\text{dg}(\lambda)$ , converting north steps to beads (1's) and east steps to gaps (0's), and using  $m$  to decide which step on the frontier corresponds to position 0 of  $w$ .
- *Jacobi Triple Product Identity.* Abaci can be used to prove

$$\sum_{m \in \mathbb{Z}} q^{m(m+1)/2} u^m = \prod_{n \geq 1} (1 + uq^n) \prod_{n \geq 0} (1 + u^{-1}q^n) \prod_{n \geq 1} (1 - q^n).$$

One consequence is the formula  $\prod_{n \geq 1} (1 - q^n) = \sum_{k \in \mathbb{Z}} (-1)^k q^{(3k^2 - k)/2}$ .

- $k$ -cores and  $k$ -quotients.* Given a partition  $\mu$ , repeated removal of border ribbons of size  $k$  (in any order) will lead to a unique partition from which no further ribbons of this kind can be removed. This partition is called the  $k$ -core of  $\mu$ . We can also find the  $k$ -core by converting  $\mu$  to an abacus, decimating the abacus to give a  $k$ -runner abacus, justifying all runners, and converting back to a partition. Each ribbon removal corresponds to moving one bead one step to the left on the  $k$ -runner abacus. Justifying each separate runner on the  $k$ -runner abacus for  $\mu$  produces the  $k$ -quotients  $(\nu^0, \dots, \nu^{k-1})$  of  $\mu$ . Alternatively,  $\text{dg}(\nu^i)$  can be found by taking the cells of  $\text{dg}(\mu)$  lying due north and due west of steps of  $k$ -content  $i$  on the frontier of  $\mu$ . We get a bijection  $\Delta_k : \text{Par} \rightarrow \text{Core}(k) \times \text{Par}^k$  by mapping  $\mu$  to its  $k$ -core and  $k$ -quotients.
- Labeled Abaci and Antisymmetric Polynomials.* A polynomial  $f$  in  $N$  variables is antisymmetric iff interchanging any two adjacent variables changes the sign of  $f$ . For each  $\mu = (\mu_1 > \mu_2 > \dots > \mu_N \geq 0)$ , the polynomial  $a_\mu(x_1, \dots, x_N) = \det \|x_j^{\mu_i}\|_{1 \leq i, j \leq N}$  is antisymmetric. Writing  $\delta(N) = (N-1, N-2, \dots, 2, 1, 0)$ , the set  $\{a_{\lambda+\delta(N)} : \lambda \in \text{Par}_N\}$  is a basis for the space  $A_N$  of antisymmetric polynomials. Division by  $a_{\delta(N)} = \prod_{1 \leq i < j \leq N} (x_i - x_j)$  gives a vector space isomorphism from  $A_N$  to  $\Lambda_N$  sending  $a_{\lambda+\delta(N)}$  to the Schur polynomial  $s_\lambda = a_{\lambda+\delta(N)}/a_{\delta(N)}$ . To model the terms in  $a_{\lambda+\delta(N)}$ , we use the  $N!$  labeled abaci consisting of beads  $1, 2, \dots, N$  (in any order) at positions given by  $\lambda + \delta(N)$ .
- Rim-Hook Tableaux.* A rim-hook tableau of shape  $\lambda/\mu$  and content  $\alpha$  is obtained by enlarging the diagram of  $\mu$  using border ribbons of lengths  $\alpha_1, \alpha_2, \dots$  (in this order) until the diagram of  $\lambda$  is obtained. The set of such tableaux is denoted  $\text{RHT}(\lambda/\mu, \alpha)$ . A ribbon occupying  $r$  rows has sign  $(-1)^{r-1}$ , and the sign of a rim-hook tableau is the product of the signs of its ribbons. We write  $\chi_\alpha^{\lambda/\mu} = \sum_{T \in \text{RHT}(\lambda/\mu, \alpha)} \text{sgn}(T)$ . We have  $\chi_\alpha^{\lambda/\mu} = \chi_\beta^{\lambda/\mu}$  whenever  $\text{sort}(\alpha) = \text{sort}(\beta)$ .
- Interactions between Abaci and Tableaux.* One can give combinatorial proofs of several identities in Table 11.1 by using the word of a tableau to encode bead motions on abaci. When these motions lead to bead collisions, one obtains two objects of opposite sign and equal weight that cancel terms on one side of the formula to be proved. Objects with no collisions are fixed points that can be reorganized to give the other side of the formula.
- Inverse Kostka Matrix.* A rim-hook tableau is called special iff each ribbon in the tableau begins in the leftmost column;  $\text{SRHT}(\mu, \lambda)$  is the set of such tableaux of shape  $\mu$  and content  $\alpha$  with  $\text{sort}(\alpha) = \lambda$ . Letting  $K'_{\lambda, \mu} = \sum_{S \in \text{SRHT}(\mu, \lambda)} \text{sgn}(S)$ , we have  $\mathbf{K}\mathbf{K}' = \mathbf{I}$ .
- Littlewood-Richardson Coefficients.* The scalars  $c_{\nu, \mu}^\lambda = c_{\mu, \nu}^\lambda$  appearing in the Schur expansions of  $s_{\lambda/\nu}$  and  $s_\mu s_\nu$  count semistandard tableaux  $T$  of shape  $\lambda/\nu$  and content  $\mu$  such that every suffix of the word of  $T$  has partition content. The scalars  $c_{\nu, \mu}^\lambda$  also count semistandard tableaux  $T$  of shape  $\mu$  and content  $\lambda - \nu$  such that  $w(T) = w_1 \dots w_n$  satisfies the following condition: for all  $k \leq n$ , the exponent vector of  $\prod_j x_j^{\nu_j} \prod_{i=k}^n x_{w_i}$  is a partition.

---

## Exercises

- 11.73.** Let  $w = \cdots 11011011\underline{10}101001100 \cdots$ . Compute  $\text{wt}(w)$  and  $J(w)$ .
- 11.74.** Compute  $U(-1, \mu)$  for each  $\mu \in \text{Par}(5)$ .
- 11.75.** In the computation of  $U(m, \mu)$  in 11.4, describe in detail how to use  $m$  to decide which symbol in the bead-gap sequence is  $w_0$ .
- 11.76.** Given  $\mu \in \text{Par}$ , what is the relationship between the abaci  $U(-1, \mu)$  and  $U(-1, \mu')$ ?
- 11.77.** Show that the abacus  $w$  in 11.2 and its justification  $J(w)$  have the same weight, if we use the weights defined in the proof of 11.5.
- 11.78.** Show how to deduce Euler's pentagonal number theorem as an algebraic consequence of the Jacobi triple product identity.
- 11.79.** Fill in all the details in the proof of 11.6.
- 11.80.** Use 11.5 to simplify the product  $\prod_{n \geq 0} (1 - x^{5n+1})^{-1} \prod_{n \geq 0} (1 - x^{5n+4})^{-1}$  appearing in one of the Rogers-Ramanujan identities. Can you give a direct proof of the resulting identity using abaci?
- 11.81.** Use 11.4 to find a bijective proof of 11.5 that makes no reference to abaci, instead using combinatorial operations on partition diagrams and their frontiers.
- 11.82.** Complete the proof of 11.12 by verifying that  $D_k(w) \in \text{Abc}^k$ ,  $I_k(v) \in \text{Abc}$ , and  $D_k$  and  $I_k$  are two-sided inverses.
- 11.83.** (a) Verify that the 3-core of  $\mu = (10, 10, 10, 8, 8, 8, 7, 4)$  is  $(1, 1)$  by removing border 3-ribbons from  $\mu$  in several different orders. (b) Use the 3-runner abacus encoding  $\mu$  to determine exactly how many ways there are to change  $\mu$  into  $(1, 1)$  by removing an ordered sequence of border 3-ribbons.
- 11.84.** Let  $\mu = (8, 7, 6, 4, 4, 4, 3, 1, 1, 1)$ . Use abaci to compute the  $k$ -core and  $k$ -quotients of  $\mu$  for  $1 \leq k \leq 6$ .
- 11.85.** Find all integer partitions that are 2-cores, and draw some of their diagrams.
- 11.86.** Find all 3-cores with at most 8 cells.
- 11.87.** Verify the assertion in the last sentence of 11.20.
- 11.88.** Let  $\mu = (8, 8, 8, 8, 8, 8, 8, 8)$ . (a) Use abaci to compute the  $k$ -core and  $k$ -quotients of  $\mu$  for  $3 \leq k \leq 8$ . (b) Use the construction at the end of §11.4 to compute the  $k$ -quotients of  $\mu$  (for  $3 \leq k \leq 8$ ) directly from the diagram of  $\mu$ .
- 11.89.** Compute the  $k$ -quotients of  $\mu = (6, 6, 6, 3, 3, 2, 2, 2, 1, 1)$  without using abaci, for  $k = 3, 4, 5$ .
- 11.90.** Consider the construction at the end of §11.4 for computing  $k$ -quotients of  $\mu$ . Show that the hook-length of each unerased cell is divisible by  $k$ , and these are the only cells in the diagram of  $\mu$  whose hook-lengths are divisible by  $k$ .
- 11.91.** For each  $k \geq 1$ , find a formula for the generating function  $\sum_{\mu \in \text{Core}(k)} q^{|\mu|}$ .

**11.92.** Given that  $\mu$  has  $k$ -core  $\rho$  and  $k$ -quotients  $\nu^0, \dots, \nu^{k-1}$ , find a formula for the number of ways we can go from  $\mu$  to  $\rho$  by removing an ordered sequence of border  $k$ -ribbons.

**11.93.** Compute  $a_\mu(x_1, x_2, x_3)$  and  $a_{\lambda+\delta(3)}(x_1, x_2, x_3)$  for  $\mu = (6, 3, 1)$  and  $\lambda = (2, 2, 1)$ .

**11.94.** Verify by direct calculation that, for  $N = 3$  and  $\lambda = (2, 1, 0)$ ,  $a_{\lambda+\delta(N)}$  is divisible by  $a_{\delta(N)}$  and  $a_{\lambda+\delta(N)}/a_{\delta(N)} = s_\lambda(x_1, \dots, x_N)$ .

**11.95.** Verify that  $A_N$  and  $A_N^n$  are subspaces of  $K[x_1, \dots, x_N]$ , and that the map  $f \mapsto fa_{\delta(N)}$  (for  $f \in \Lambda_N$ ) is  $K$ -linear.

**11.96.** (a) Show that the product of two antisymmetric polynomials is symmetric. (b) Show that the product of a symmetric polynomial and an antisymmetric polynomial is antisymmetric.

**11.97.** Define a map  $T : K[x_1, \dots, x_N] \rightarrow K[x_1, \dots, x_N]$  by setting

$$T(f) = \frac{1}{N!} \sum_{w \in S_N} \text{sgn}(w) f(x_{w(1)}, \dots, x_{w(N)}).$$

Show that  $T$  is a  $K$ -linear map with image  $A_N$  whose restriction to  $A_N$  is the identity map. Can you describe the kernel of  $T$ ?

**11.98.** Let  $v$  be the labeled abacus  $v = 0041000300502600 \dots$ . Compute  $\text{wt}(v)$ ,  $w(v)$ ,  $\text{pos}(v)$ , and  $\text{sgn}(v)$ . For which  $\lambda$  is  $v$  in  $\text{LAbc}(\lambda + \delta(6))$ ?

**11.99.** Draw all the labeled abaci in  $\text{LAbc}(6, 5, 1)$ , and compute the sign of each abacus.

**11.100.** Using  $N = 6$  variables, compute:

(a)  $a_{(4,2,1)+\delta(6)}p_4$ ; (b)  $a_{(3,3,3)+\delta(6)}p_3$ ; (c)  $a_{(1,1,1,1,1)+\delta(6)}p_2$ .

How would the answers change if we changed  $N$ ?

**11.101.** Let  $v = 0310040206500 \dots \in \text{LAbc}(\lambda + \delta(6))$  and  $k = 4$ . For  $1 \leq i \leq 6$ , compute  $I(v, i)$  where  $I$  is the involution in the proof of 11.39. For any fixed points that arise, compute  $v^*$  and indicate which border 4-ribbon is added to  $\text{dg}(\lambda)$  in the passage from  $v$  to  $v^*$ .

**11.102.** Using  $N = 6$  variables, compute:

(a)  $a_{(4,2,1)+\delta(6)}e_3$ ; (b)  $a_{(3,3,3)+\delta(6)}e_2$ ; (c)  $a_{(5,4,3,1,1)+\delta(6)}e_4$ .

How would the answers change if we changed  $N$ ?

**11.103.** Let  $v = 0310040206500 \dots \in \text{LAbc}(\lambda + \delta(6))$ . Compute  $I(v, S)$  for  $S = \{2, 5, 6\}$ ,  $S = \{1, 4, 5\}$ ,  $S = \{1, 3, 4\}$ , and  $S = \{3, 4, 6\}$ , where  $I$  is the involution in the proof of 11.42. For any fixed points that arise, compute  $v^*$  and indicate which vertical strip is added to  $\text{dg}(\lambda)$  in the passage from  $v$  to  $v^*$ .

**11.104.** Using  $N = 5$  variables, compute:

(a)  $a_{(4,2,1)+\delta(6)}h_3$ ; (b)  $a_{(3,3,3)+\delta(6)}h_3$ ; (c)  $a_{(5,4,3,1,1)+\delta(6)}h_4$ .

How would the answers change if we changed  $N$ ?

**11.105.** Let  $v = 0310040206500 \dots \in \text{LAbc}(\lambda + \delta(6))$ . Compute  $I(v, M)$  for  $M = [1, 1, 4, 5]$ ,  $M = [2, 2, 5, 6]$ ,  $M = [2, 4, 5, 5]$ , and  $M = [1, 2, 3, 4]$ , where  $I$  is the involution in the proof of 11.44. For any fixed points that arise, compute  $v^*$  and indicate which horizontal strip is added to  $\text{dg}(\lambda)$  in the passage from  $v$  to  $v^*$ .

**11.106.** Explain in detail why the bead motion rule in §11.9 leads to the addition of a horizontal  $k$ -strip to the shape  $\lambda$ , assuming no bead collision occurs.

**11.107.** In the proof of 11.44, check in detail that  $I$  reverses signs, preserves weights, and is an involution.

**11.108.** Reprove 11.45 by comparing the symmetric and antisymmetric Pieri rules for multiplication by  $e_k$ .

**11.109.** Expand the following symmetric polynomials into linear combinations of Schur polynomials: (a)  $s_{(3,3,2)}p_3$ ; (b)  $p_{(3,1,3)}$ ; (c)  $s_{(2,2)}p_{(2,1)}$ .

**11.110.** Compute the coefficients of the following Schur polynomials in the Schur expansion of  $p_{(3,3,2,1)}$ : (a)  $s_{(9)}$ ; (b)  $s_{(3,3,3)}$ ; (c)  $s_{(4,4,1)}$ ; (d)  $s_{(1^9)}$ .

**11.111.** Show that, for  $\lambda \in \text{Par}(n)$ ,  $\chi_{(1^n)}^\lambda = |\text{SYT}(\lambda)|$ .

**11.112.** Write  $s_{(3,2,1)}$  as a linear combination of power-sum polynomials.

**11.113.** For each  $\mu \in \text{Par}(4)$ , write  $p_\mu$  in terms of Schur polynomials, and write  $s_\mu$  in terms of power-sum polynomials.

**11.114.** Let  $I$  be the involution in §11.12. For each  $(v, T) \in X$  given below, compute  $I(v, T)$ . If  $(v, T)$  is a fixed point, compute  $v^* \in Y$ .

(a)  $v = 5432100 \dots$ ,  $T =$ 

1	1	1	2	2
2	3	4		
3	5			

.

(b)  $v = 2431500 \dots$ ,  $T =$ 

5	5	5	5	5
1	1	1		
3	3			

.

(c)  $v = 3452100 \dots$ ,  $T =$ 

1	1	1	1	1
2	2	4		
5	3			

.

**11.115.** Let  $I$ ,  $X$ , and  $Y$  be defined as in §11.12. Take  $N = 3$  and  $\lambda = (2, 1, 0)$ . List all the elements of  $X$  and  $Y$ , compute the action of  $I$  on  $X$ , and show how the fixed points of  $I$  map bijectively to  $Y$ .

**11.116.** Verify all the assertions stated just before 11.54.

**11.117.** Express  $s_{(4,3,1)/(2,1)}$  as a linear combination of power-sums.

**11.118.** Explain why the formulas  $\omega(h_\mu) = e_\mu$  and  $\omega(e_\mu) = h_\mu$  are special cases of 11.59.

**11.119.** For  $N \geq k$ , two linear operators  $S$  and  $T$  on  $\Lambda_N^k$  are called *adjoint* iff  $\langle S(f), g \rangle = \langle f, T(g) \rangle$  for all  $f, g \in \Lambda_N^k$ . Prove that this condition holds for all such  $f, g$  iff it holds for all  $f$  in some basis of  $\Lambda_N^k$  and all  $g$  in some (possibly different) basis of  $\Lambda_N^k$ .

**11.120.** Write the following Schur polynomials in terms of the complete symmetric polynomials  $h_\mu$ : (a)  $s_{(5,3)}$ ; (b)  $s_{(4,1,1)}$ ; (c)  $s_{(5,5,2,2)}$ .

**11.121.** Write the following Schur polynomials in terms of the elementary symmetric polynomials  $e_\mu$ : (a)  $s_{(2,2,2,2)}$ ; (b)  $s_{(3,2,1)}$ ; (c)  $s_{(4,2)}$ .

**11.122.** Write the skew Schur polynomial  $s_{(4,4,3)/(2,1,1)}$  in terms of: (a) the  $h_\mu$ 's; (b) the  $e_\mu$ 's; (c) the  $p_\mu$ 's; (d) the  $m_\mu$ 's.

**11.123.** Modify the definition of the involution used in the proof of 11.60 as follows. If two or more paths in  $(w, p_1, \dots, p_n)$  intersect, choose  $i$  minimal and then  $j$  minimal such that  $p_i$  and  $p_j$  intersect. Let  $(u, v)$  be the earliest vertex on  $p_i$  that is also a vertex of  $p_j$ , and switch the initial segments of these two paths as in the original proof. Show that the map just defined is *not* always an involution.

**11.124.** Can you find a way to rephrase the combinatorial proof of 11.60 in terms of abaci?

**11.125.** Enumerate special rim-hook tableaux to compute  $K'_{\mu,\lambda}$  for all partitions  $\lambda, \mu$  with at most 4 cells. Use this to confirm by direct calculation that  $\mathbf{K}\mathbf{K}' = \mathbf{I}$ .

**11.126.** Find and prove a Pieri-type rule giving the Schur expansion of a product  $s_\nu m_\lambda$ .

**11.127.** Let  $\mathbf{K}'$  be the matrix defined combinatorially by  $K'_{\lambda,\mu} = \sum_{S \in \text{SRHT}(\mu,\lambda)} \text{sgn}(S)$ . Find involutions that prove  $\mathbf{K}\mathbf{K}' = \mathbf{I}$ .

**11.128.** Let  $\mathbf{K}'$  be the inverse Kostka matrix, defined using special rim-hook tableaux. Can you prove the identity  $\mathbf{K}'\mathbf{K} = \mathbf{I}$  combinatorially?

**11.129.** Let  $I$  be the involution in the proof of 11.66. (a) Compute  $I(v^0, T)$ , where

$$v_0 = 5432100 \cdots, \quad T = \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 2 & 2 & 3 \\ \hline 1 & 2 & 3 & 4 & 4 \\ \hline 2 & 3 & 5 \\ \hline \end{array}.$$

(b) Answer (a) if the last 1 in the top row of  $T$  is changed to a 2. (c) Answer (a) if the last 3 in row 2 of  $T$  is changed to a 2.

**11.130.** Compute  $c_{\nu,\mu}^\lambda$  and  $c_{\mu,\nu}^\lambda$  using 11.66, where: (a)  $\lambda = (5, 3, 1, 1)$ ,  $\mu = (3, 1)$ ,  $\nu = (3, 2, 1)$ ; (b)  $\lambda = (5, 4, 4, 3, 1)$ ,  $\mu = (4, 3, 3, 1)$ ,  $\nu = (3, 1, 1, 1)$ .

**11.131.** Repeat the previous exercise, but use 11.72 to compute the Littlewood-Richardson coefficients.

**11.132.** Continuing 11.69, find the expansion of  $s_{(5,4,4,1)/(3,1)}$  into a sum of Schur polynomials.

**11.133.** Expand the following skew Schur polynomials into sums of Schur polynomials: (a)  $s_{(3,3,3)/(2,1)}$ ; (b)  $s_{(5,4)/(2)}$ ; (c)  $s_{(4,3,2,1)/(1,1,1)}$ .

**11.134.** Expand  $s_{(3,2)}s_{(2,2)}$  into a sum of Schur polynomials.

**11.135.** In the Schur expansion of  $s_{(3,2,1,1)}^2$ , find the coefficients of: (a)  $s_{(5,4,2,2,1)}$ ; (b)  $s_{(5,3,3,1,1,1)}$ ; (c)  $s_{(4,3,3,2,1,1)}$ .

**11.136.** Give a combinatorial proof of 11.72 based on abaci.

## Notes

The proof of the Jacobi triple product identity in §11.2 is adapted from a lecture of Richard Borcherds. One source for material on unlabeled abaci,  $k$ -cores, and  $k$ -quotients is the book by James and Kerber [72]; for labeled abaci, see Loehr [84]. Gessel and Viennot [53] have used intersecting lattice path models to prove many enumeration results. The combinatorial interpretation of the inverse Kostka matrix is due to Eğecioğlu and Remmel [33]. The proof of the Littlewood-Richardson rule given in §11.16 may be viewed as a combinatorialization of the algebraic proof in Remmel and Shimozono [111]. Many other proofs of this rule may be found in the literature; see, e.g., the bibliographic notes in Fulton [46] and Stanley [127, Ch. 7].

## Additional Topics

This chapter covers a variety of topics illustrating different aspects of enumerative combinatorics and probability. The treatment of each topic is essentially self-contained.

### 12.1 Cyclic Shifting of Paths

This section illustrates another technique for enumerating certain collections of lattice paths. The basic idea is to introduce an equivalence relation on paths by cyclically shifting the steps of a path. A similar idea was used in §3.14 to enumerate lists of terms.

**12.1. Theorem: Enumeration of Rational-Slope Dyck Paths.** Let  $r$  and  $s$  be positive integers such that  $\gcd(r, s) = 1$ . The number of lattice paths from  $(0, 0)$  to  $(r, s)$  that never go below the diagonal line  $sx = ry$  is

$$\frac{1}{r+s} \binom{r+s}{r, s}.$$

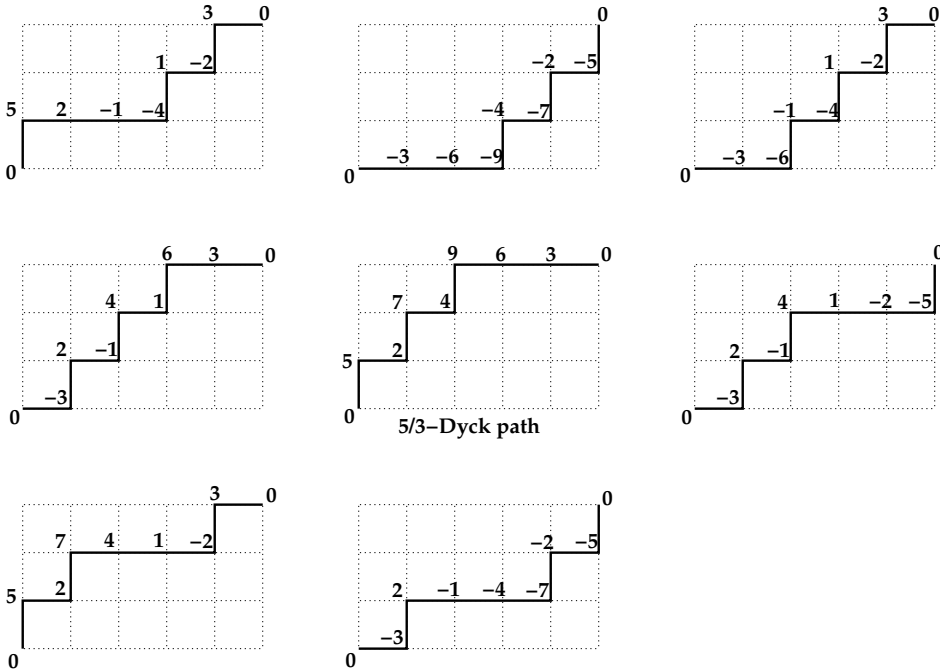
(Such paths are called  $r/s$ -Dyck paths.)

*Proof. Step 1.* Let  $X = \mathcal{R}(E^r N^s)$ , which is the set of all rearrangements of  $r$  copies of  $E$  and  $s$  copies of  $N$ . Thinking of  $E$  as an east step and  $N$  as a north step, we see that  $X$  can be identified with the set of all lattice paths from  $(0, 0)$  to  $(r, s)$ . Given  $v = v_1 v_2 \cdots v_{r+s} \in X$ , we define an associated *label vector*  $L(v) = (m_0, m_1, \dots, m_{r+s})$  as follows. We set  $m_0 = 0$ . Then we recursively calculate  $m_i = m_{i-1} + r$  if  $v_i = N$ ,  $m_i = m_{i-1} - s$  if  $v_i = E$ . For example, if  $r = 5$ ,  $s = 3$ , and  $v = \text{NEEENENE}$ , then  $L(v) = (0, 5, 2, -1, -4, 1, -2, 3, 0)$ . We can also describe this construction in terms of the lattice path encoded by  $v$ . If we label each lattice point  $(x, y)$  on this path by the integer  $ry - sx$ , then  $L(v)$  is the sequence of labels encountered as we traverse the path from  $(0, 0)$  to  $(r, s)$ . This construction is illustrated by the lattice paths in Figure 12.1. Note that  $v$  is recoverable from  $L(v)$ , since  $v_i = N$  iff  $m_i - m_{i-1} = r$  and  $v_i = E$  iff  $m_i - m_{i-1} = -s$ .

*Step 2.* We prove that for all  $v \in X$ , if  $L(v) = (m_0, m_1, \dots, m_{r+s})$  then  $m_0, m_1, \dots, m_{r+s-1}$  are *distinct*, whereas  $m_{r+s} = 0 = m_0$ . To see this, suppose there exist  $x, y, a, b$  with  $0 < a \leq r$  and  $0 < b \leq s$ , such that  $(x, y)$  and  $(x + a, y + b)$  are two points on the lattice path for  $v$  that have the same label. This means that  $ry - sx = r(y + b) - s(x + a)$ , which simplifies to  $rb = sa$ . Thus the number  $rb = sa$  is a common multiple of  $r$  and  $s$ . Since  $\gcd(r, s) = 1$ , we have  $\text{lcm}(r, s) = rs$ , so that  $rb \geq rs$  and  $sa \geq rs$ . Thus  $b \geq s$  and  $a \geq r$ , forcing  $b = s$  and  $a = r$ . But then  $(x, y)$  must be  $(0, 0)$  and  $(x + a, y + b)$  must be  $(r, s)$ . So the only two points on the path with equal labels are  $(0, 0)$  and  $(r, s)$ , which correspond to  $m_0$  and  $m_{r+s}$ .

*Step 3.* Introduce an equivalence relation  $\sim$  on  $X$  by setting  $v \sim w$  iff  $v$  is a cyclic shift of  $w$ . More precisely, defining  $C(w_1 w_2 \cdots w_{r+s}) = w_2 w_3 \cdots w_{r+s} w_1$ , we have  $v \sim w$  iff



**FIGURE 12.1**

Cyclic shifts of a lattice path.

$v = C^i(w)$  for some integer  $i$  (which can be chosen in the range  $0 \leq i < r + s$ ). For each  $v \in X$ , let  $[v] = \{w \in X : w \sim v\}$  be the equivalence class of  $v$  relative to this equivalence relation. Figure 12.1 shows the paths in the class [NEEENENE].

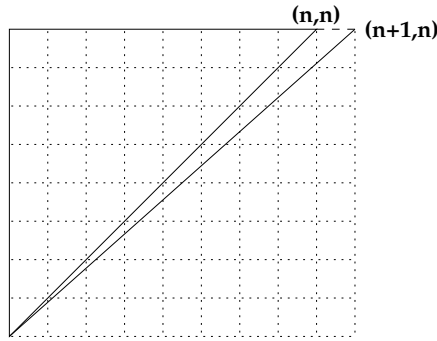
*Step 4.* We show that  $|[v]| = r + s$  for all  $v \in X$ , which means that all  $r + s$  cyclic shifts of  $v$  are *distinct*. Suppose  $v = v_1 v_2 \cdots v_{r+s}$  has  $L(v) = (m_0, m_1, \dots, m_{r+s})$ . By definition of  $L$ , for each  $i$  with  $0 \leq i < r + s$ , the label vector of the cyclic shift  $C^i(v) = v_{i+1} \cdots v_{r+s} v_1 \cdots v_i$  is

$$L(C^i(v)) = (0, m_{i+1} - m_i, m_{i+2} - m_i, \dots, m_{r+s} - m_i, m_1 - m_i, \dots, m_i - m_i)$$

(cf. Figure 12.1). The set of integers appearing in the label vector  $L(C^i(v))$  is therefore obtained from the set of integers in  $L(v)$  by subtracting  $m_i$  from each integer in the latter set. In particular, if  $\mu$  is the smallest integer in  $L(v)$ , then the smallest integer in  $L(C^i(v))$  is  $\mu - m_i$ . Since the numbers  $m_0, m_1, \dots, m_{r+s-1}$  are distinct (by step 2), we see that the minimum elements in the sequences  $L(C^i(v))$  are distinct, as  $i$  ranges from 0 to  $r + s - 1$ . This implies that the sequences  $L(C^i(v))$ , and hence the words  $C^i(v)$ , are pairwise distinct.

*Step 5.* We show that, for all  $v \in X$ , there exists a unique word  $w \in [v]$  such that  $w$  encodes a rational-slope Dyck path. By the way we defined the labels,  $w$  is an  $r/s$ -Dyck path iff  $L(w)$  has no negative entries. Recall from step 4 that the set of labels in  $L(C^i(v))$  is obtained from the set of labels in  $L(v)$  by subtracting  $m_i$  from each label in the latter set. By step 2, there is a unique  $i$  in the range  $0 \leq i < r + s$  such that  $m_i = \mu$ , the minimum value in  $L(v)$ . For this choice of  $i$ , we have  $m_j \geq \mu = m_i$  for every  $j$ , so that  $m_j - m_i \geq 0$  and  $L(C^i(v))$  has no negative labels. For any other choice of  $i$ ,  $m_i > \mu$  by step 4, so that  $L(C^i(v))$  contains the negative label  $\mu - m_i$ .

*Step 6.* Suppose  $\sim$  has  $n$  equivalence classes in  $X$ . By step 5,  $n$  is also the number of rational-slope Dyck paths. By step 4, each equivalence class has size  $r + s$ . Since  $X$  is the



**FIGURE 12.2**

Comparing Dyck paths to  $(n+1)/n$ -Dyck paths.

disjoint union of its equivalence classes, the sum rule and 1.46 give

$$\binom{r+s}{r,s} = |X| = n(r+s).$$

Dividing by  $r+s$  gives the formula stated in the theorem.  $\square$

**12.2. Corollary: Enumeration of Dyck Paths and  $m$ -Dyck Paths.** For  $n \geq 1$ , the number of Dyck paths ending at  $(n, n)$  is

$$\frac{1}{2n+1} \binom{2n+1}{n+1, n}.$$

For  $m, n \geq 1$ , the number of  $m$ -Dyck paths ending at  $(mn, n)$  is

$$\frac{1}{(m+1)n+1} \binom{(m+1)n+1}{mn+1, n}.$$

*Proof.* Let  $X$  be the set of Dyck paths ending at  $(n, n)$ , and let  $X'$  be the set of  $(n+1)/n$ -Dyck paths ending at  $(n+1, n)$ . Since  $\gcd(n+1, n) = 1$ , we know that  $|X'| = \frac{1}{2n+1} \binom{2n+1}{n+1, n}$ . On the other hand, passing from the diagonal  $y = x$  to the line  $(n+1)y - nx = 0$  does not introduce any new lattice points in the region of interest, except for  $(n+1, n)$ . See Figure 12.2. It follows that appending a final east step gives a bijection from  $X$  onto  $X'$ , so the first result holds. The second result is proved in the same way: appending a final east step gives a bijection from the set of  $m$ -Dyck paths ending at  $(mn, n)$  to the set of  $(mn+1)/n$ -Dyck paths ending at  $(mn+1, n)$ .  $\square$

## 12.2 Chung-Feller Theorem

In §1.10, we defined Dyck paths and proved that the number of Dyck paths of order  $n$  is the Catalan number  $C_n = \frac{1}{n+1} \binom{2n}{n, n}$ . This section discusses a remarkable generalization of this result called the *Chung-Feller Theorem*.

**12.3. Definition: Flawed Paths.** Suppose  $\pi = ((x_0, y_0), \dots, (x_{2n}, y_{2n}))$  is a lattice path from  $(0, 0)$  to  $(n, n)$ . For  $1 \leq j \leq n$ , we say that  $\pi$  has a *flaw in row  $j$*  iff there exists a point  $(x_i, y_i)$  visited by  $\pi$  such that  $y_i = j - 1$ ,  $y_i < x_i$ , and  $(x_{i+1}, y_{i+1}) = (x_i, y_i + 1)$ . This means that the  $j$ th north step of  $\pi$  occurs in the region southeast of the diagonal line  $y = x$ . For  $1 \leq j \leq n$ , define

$$X_j(\pi) = \chi(\pi \text{ has a flaw in row } j).$$

Also define the *number of flaws of  $\pi$*  by setting  $\text{flaw}(\pi) = X_1(\pi) + X_2(\pi) + \dots + X_n(\pi)$ .

For example, the paths shown in Figure 12.3 have zero and six flaws, respectively. The paths shown in Figure 12.4 have five and zero flaws, respectively. Observe that  $\pi$  is a *Dyck* path iff  $\text{flaw}(\pi) = 0$ .

**12.4. Chung-Feller Theorem.** Fix  $n \geq 0$ , and let  $A$  be the set of lattice paths from  $(0, 0)$  to  $(n, n)$ . For  $0 \leq k \leq n$ , let

$$A_k = \{\pi \in A : \text{flaw}(\pi) = k\}.$$

Then  $|A_k| = |A_0|$  for all  $k$ . In particular, for  $0 \leq k \leq n$ ,

$$|A_k| = \frac{1}{n+1}|A| = \frac{1}{n+1} \binom{2n}{n, n} = C_n.$$

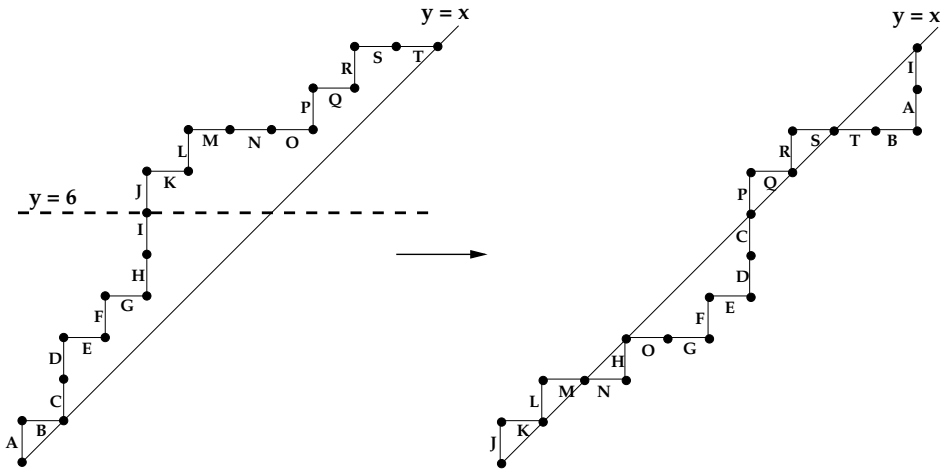
*Proof.* Fix  $k > 0$ . To prove that  $|A_0| = |A_k|$ , we define a bijection  $\phi_k : A_0 \rightarrow A_k$ . See Figure 12.3 for an example where  $n = 10$  and  $k = 6$ . Given a Dyck path  $\pi \in A_0$ , we begin by drawing the line  $y = k$  superimposed on the Dyck path. There is a unique point  $(x_i, y_i)$  on  $\pi$  such that  $y_i = k$  and  $\pi$  arrives at  $(x_i, y_i)$  by taking a vertical step. Call this step the *special vertical step*. Let  $(a_1, a_2, \dots) \in \{H, V\}^{2n-x_i-y_i}$  be the ordered sequence of steps of  $\pi$  reading northeast from  $(x_i, y_i)$ , where H means “horizontal step” and V means “vertical step.” Let  $(b_0 = V, b_1, b_2, \dots) \in \{H, V\}^{x_i+y_i}$  be the ordered sequence of steps of  $\pi$  reading southwest from  $(x_i, y_i)$ . For the Dyck path shown on the left in Figure 12.3, we have

$$a_1 a_2 \dots = \text{VHVHHHVHVHH}, \quad b_0 b_1 b_2 \dots = \text{VHVHVHVHV}.$$

We compute  $\phi_k(\pi) = c_1 c_2 \dots c_{2n} \in \{V, H\}^{2n}$  as follows. Let  $c_1 = a_1$ ,  $c_2 = a_2$ , etc., until we obtain a horizontal step  $c_k (= a_k)$  that ends strictly below the diagonal  $y = x$ . Then set  $c_{k+1} = b_1$ ,  $c_{k+2} = b_2$ , etc., until we obtain a vertical step  $c_{k+m} (= b_m)$  that ends on the line  $y = x$ . Then set  $c_{k+m+1} = a_{k+1}$ ,  $c_{k+m+2} = a_{k+2}$ , etc., until we take a horizontal step that ends strictly below  $y = x$ . Then switch back to using the steps  $b_{m+1}, \dots$  until we return to  $y = x$ . Continue in this way until all steps are used. By convention, the special vertical step  $b_0 = V$  is the last “ $b$ -step” to be consumed.

For example, for the path  $\pi$  in Figure 12.3, we have labeled the steps of  $\pi$  as A through T for ease of reference. The special vertical step is step I. We begin by transferring steps J, K, L, M, N to the image path (starting at the origin). Step N goes below the diagonal, so we jump to the section of  $\pi$  prior to the special vertical step and work southwest. After taking only one step (step H), we have returned to the diagonal. Now we jump back to our previous location in the top part of  $\pi$  and take step O. This again takes us below the diagonal, so we jump back to the bottom part of  $\pi$  and transfer steps G, F, E, D, C. Now we return to the top part and transfer steps P, Q, R, S, T. Finally, we return to the bottom part of  $\pi$  and transfer steps B, A, and finally the special vertical step I.

This construction has the following crucial property. Vertical steps above the line  $y = k$  in  $\pi$  get transferred to vertical steps above the line  $y = x$  in  $\phi_k(\pi)$ , while vertical steps below the line  $y = k$  in  $\pi$  get transferred to vertical steps below the line  $y = x$  in  $\phi_k(\pi)$ . Thus,  $\phi_k(\pi)$  has exactly  $k$  flaws, and is therefore an element of  $A_k$ .



**FIGURE 12.3**

Mapping Dyck paths to flawed paths.

Moreover, consider the coordinates of the special point  $(x_i, y_i)$ . By definition,  $y_i = k = \text{flaw}(\phi_k(\pi))$ . On the other hand, we claim that  $y_i - x_i$  equals the number of horizontal steps in  $\phi_k(\pi)$  that start on  $y = x$  and end to the right of  $y = x$ . Each such horizontal step corresponds to a step after  $(x_i, y_i)$  in  $\pi$  that brings the path closer to the main diagonal  $y = x$ . For instance, these steps are N, O, and T in Figure 12.3. The definition of  $\phi_k$  shows that the steps in question (in  $\pi$ ) are the earliest east steps after  $(x_i, y_i)$  that arrive on the lines  $y = x + d$  for  $d = y_i - x_i - 1, \dots, 2, 1, 0$ . The number of such steps is therefore  $y_i - x_i$  as claimed.

The observations in the last paragraph allow us to compute the inverse map  $\phi'_k : A_k \rightarrow A_0$ . For, suppose  $\pi \in A_k$  is a path with  $k$  flaws. We can recover  $(x_i, y_i)$  since  $y_i = k$  and  $y_i - x_i$  is the number of east steps of  $\pi$  departing from  $y = x$ . Next, we transfer the steps of  $\pi$  to the top and bottom portions of  $\phi'_k(\pi)$  by reversing the process described earlier. Figure 12.4 gives an example where  $n = 10$  and  $k = 5$ . First we find the special point  $(x_i, y_i) = (2, 5)$ . We start by transferring the initial steps A,B,C of  $\pi$  to the part of the image path starting at  $(2, 5)$  and moving northeast. Since C goes below the diagonal in  $\pi$ , we now switch to the bottom part of the image path. The special vertical step must be skipped, so we work southwest from  $(2, 4)$ . We transfer steps D,E,F,G,H. Since H returns to  $y = x$  in  $\pi$ , we then switch back to the top part of the image path. We only get to transfer one step (step I) before returning to the bottom part of the image path. We transfer step J, then move back to the top part and transfer steps K through S. Finally, step T is transferred to become the special vertical step from  $(2, 4)$  to  $(2, 5)$ . One checks that  $\phi'_k$  is the two-sided inverse of  $\phi_k$ , so  $\phi_k : A_0 \rightarrow A_k$  is a bijection.

Now that we know  $|A_k| = |A_0|$  for all  $k$ , the final statement of the theorem follows. For  $A$  is the disjoint union of the  $n + 1$  sets  $A_0, A_1, \dots, A_n$ , all of which have cardinality  $|A_0|$ . By the sum rule,

$$|A| = |A_0| + |A_1| + \dots + |A_n| = (n + 1)|A_0|,$$

and therefore

$$|A_k| = |A_0| = \frac{|A|}{n + 1} = \frac{1}{n + 1} \binom{2n}{n, n} = C_n$$

for  $0 \leq k \leq n$ . □

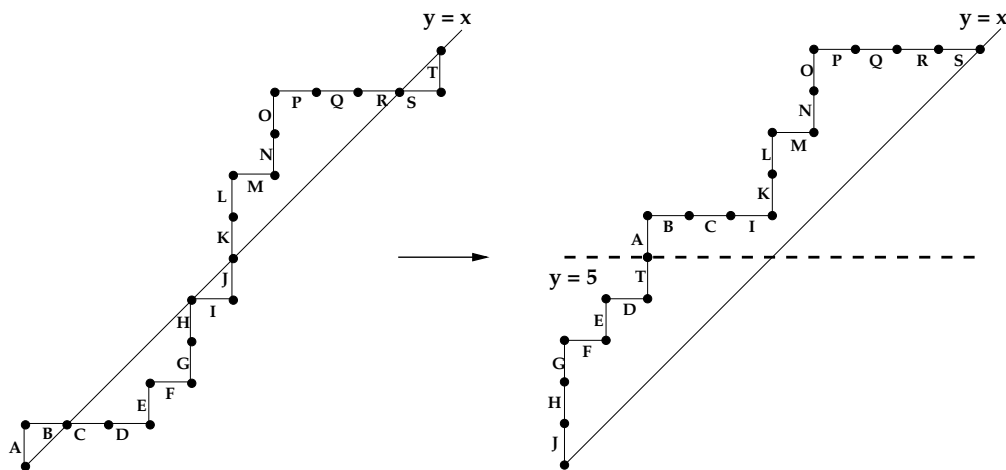


FIGURE 12.4

Mapping flawed paths to Dyck paths.

In probabilistic language, the Chung-Feller Theorem can be stated as follows.

**12.5. Corollary.** Suppose we pick a random lattice path  $\pi$  from the origin to  $(n, n)$ . The number of flaws in this path is uniformly distributed on  $\{0, 1, 2, \dots, n\}$ . In other words,

$$P(\text{flaw}(\pi) = k) = \frac{1}{n+1} \quad \text{for } k = 0, 1, \dots, n.$$

*Proof.* We compute

$$P(\text{flaw}(\pi) = k) = \frac{|A_k|}{|A|} = \frac{\frac{1}{n+1} \binom{2n}{n,n}}{\binom{2n}{n,n}} = \frac{1}{n+1}. \quad \square$$

**12.6. Remark.** The Chung-Feller Theorem is significant in probability theory for the following reason. One of the most celebrated theorems of probability is the *central limit theorem*. Roughly speaking, this theorem says that the sum of a large number of independent, identically distributed random variables (suitably normalized) will converge to a *normal distribution*. The normal distribution is described by the “bell curve” that appears ubiquitously in probability and statistics. One often deals with situations involving random variables that are *not* identically distributed and are *not* independent of one another. One might hope that a generalization of the central limit theorem would still hold in such situations.

Chung and Feller used the example of flawed lattice paths to show that such a generalization is not always possible. Fix  $n > 0$ , and let the sample space  $S$  consist of all lattice paths from the origin to  $(n, n)$ . Given a lattice path  $\pi \in S$ , recall that

$$\text{flaw}(\pi) = X_1(\pi) + X_2(\pi) + \dots + X_n(\pi),$$

where  $X_j(\pi) = \chi(\pi \text{ has a flaw in row } j)$ . The random variables  $X_1, X_2, \dots, X_n$  are identically distributed; in fact, one can show that  $P(X_j = 0) = 1/2 = P(X_j = 1)$  for all  $j$  (see 12.96). But we have seen that the sum of these random variables, namely  $X_1 + X_2 + \dots + X_n = \text{flaw}$ , is uniformly distributed on  $\{0, 1, 2, \dots, n\}$  for every  $n$ . A uniform distribution is about as far as one can get from a normal distribution! The trouble is that the random variables  $X_1, \dots, X_n$  are not independent.

## 12.3 Rook-Equivalence of Ferrers Boards

This section continues the investigation of rook theory begun in §2.11. We define the notion of a rook polynomial for a Ferrers board and derive a characterization of when two Ferrers boards have the same rook polynomial.

**12.7. Definition: Ferrers Boards and Rook Polynomials.** Let  $\mu = (\mu_1 \geq \mu_2 \geq \cdots \geq \mu_s > 0)$  be an integer partition of  $n$ . The *Ferrers board*  $F_\mu$  is a diagram consisting of  $s$  left-justified rows of squares with  $\mu_i$  squares in row  $i$ . A *non-attacking placement of  $k$  rooks* on  $F_\mu$  is a subset of  $k$  squares in  $F_\mu$  such that no two squares lie in the same row or column. Let  $r_k(\mu)$  be the number of non-attacking placements of  $k$  rooks on  $F_\mu$ . The *rook polynomial* of  $\mu$  is

$$R_\mu(x) = \sum_{k \geq 0} r_k(\mu) x^k.$$

**12.8. Example.** If  $\mu = (4, 1, 1, 1)$ , then  $R_\mu(x) = 9x^2 + 7x + 1$ . To see this, note that there is one empty subset of  $F_\mu$  (which is a non-attacking placement of zero rooks). We can place one rook on any of the 7 squares in  $F_\mu$ , so the coefficient of  $x^1$  in  $R_\mu$  is 7. To place two non-attacking rooks, we place one rook in the first column but not in the first row (3 ways), and we place the second rook in the first row but not in the first column (3 ways). The product rule gives 9 as the coefficient of  $x^2$  in  $R_\mu$ . It is impossible to place three or more non-attacking rooks on  $F_\mu$ , so all higher coefficients in  $R_\mu$  are zero.

As seen in the previous example, the constant term in any rook polynomial is 1, whereas the linear coefficient of a rook polynomial is the number  $|\mu|$  of squares on the board  $F_\mu$ . Furthermore,  $R_\mu(x)$  has degree at most  $\min(\mu_1, \ell(\mu))$ , since all rooks must be placed in distinct rows and columns of the board.

It is possible for two different partitions to have the same rook polynomial. For example, one may check that

$$R_{(2,2)}(x) = 2x^2 + 4x + 1 = R_{(3,1)}(x) = R_{(2,1,1)}(x).$$

More generally,  $R_\mu(x) = R_{\mu'}(x)$  for any partition  $\mu$ .

**12.9. Definition: Rook-Equivalence.** We say that two integer partitions  $\mu$  and  $\nu$  are *rook-equivalent* iff they have the same rook polynomial, which means  $r_k(\mu) = r_k(\nu)$  for all  $k \geq 0$ .

A necessary condition for  $\mu$  and  $\nu$  to be rook equivalent is that  $|\mu| = |\nu|$ . The next theorem gives an easily tested necessary and sufficient criterion for deciding whether two partitions are rook-equivalent.

**12.10. Theorem: Rook-Equivalence of Ferrers Boards.** Suppose  $\mu$  and  $\nu$  are partitions of  $n$ . Write  $\mu = (\mu_1 \geq \cdots \geq \mu_n)$  and  $\nu = (\nu_1 \geq \cdots \geq \nu_n)$  by adding zero parts if necessary. The rook polynomials  $R_\mu(x)$  and  $R_\nu(x)$  are equal iff the multisets

$$[\mu_1 + 1, \mu_2 + 2, \dots, \mu_n + n] \text{ and } [\nu_1 + 1, \nu_2 + 2, \dots, \nu_n + n]$$

are equal.

*Proof.* The idea of the proof is to use the falling factorial basis  $\{(x)_{\downarrow n} : n \geq 0\}$  for the vector

space of polynomials in  $x$  instead of the monomial basis  $\{x^n : n \geq 0\}$  (see 2.76). For any partition  $\lambda$ , define

$$R'_\lambda(x) = \sum_{k=0}^n r_{n-k}(\lambda)(x) \downarrow_k = \sum_{k=0}^n r_{n-k}(\lambda)x(x-1)\cdots(x-k+1).$$

Note that  $R_\mu(x) = R_\nu(x)$  iff  $r_k(\mu) = r_k(\nu)$  for  $0 \leq k \leq n$  (by linear independence of the monomial basis) iff  $R'_\mu(x) = R'_\nu(x)$  in  $\mathbb{R}[x]$  (by linear independence of the falling factorial basis). We will prove that this last condition holds iff the multisets mentioned in the theorem are equal.

We now use rook combinatorics to derive a formula for  $R'_\mu(x)$ . Fix a positive integer  $x$ . Consider the extended board  $F_\mu(x)$ , which has  $\mu_i + x$  squares in row  $i$ , for  $1 \leq i \leq n$ . We obtain  $F_\mu(x)$  from the board  $F_\mu$  by adding  $x$  new squares on the left end of each of the  $n$  rows. Let us count the number of placements of  $n$  non-attacking rooks on  $F_\mu(x)$ . On one hand, we can build such a placement by working up the rows from bottom to top, placing a rook in a valid column of each successive row. By the product rule, the number of valid placements is

$$(x + \mu_n)(x + \mu_{n-1} - 1) \cdots (x + \mu_1 - (n-1)) = \prod_{i=1}^n (x + [\mu_i - (n-i)]).$$

On the other hand, let us count the number of placements of  $n$  non-attacking rooks on  $F_\mu(x)$  that have exactly  $k$  rooks on the original board  $F_\mu$ . We can place these rooks first in  $r_k(\mu)$  ways. The remaining  $n-k$  rooks must go in the remaining  $n-k$  unused rows in one of the leftmost  $x$  squares. Placing these rooks one at a time, we obtain  $r_k(\mu)x(x-1)(x-2)\cdots(x-(n-k-1))$  valid placements. Adding over  $k$  gives the identity

$$\sum_{k=0}^n r_k(\mu)(x) \downarrow_{n-k} = \prod_{i=1}^n (x + [\mu_i - (n-i)]).$$

Replacing  $k$  by  $n-k$  in the summation, we find that

$$R'_\mu(x) = \prod_{i=1}^n (x + [\mu_i - (n-i)]).$$

This polynomial identity holds for infinitely many values of  $x$  (namely, for each positive integer  $x$ ), so the identity must hold in the polynomial ring  $\mathbb{R}[x]$ . Similarly,

$$R'_\nu(x) = \prod_{i=1}^n (x + [\nu_i - (n-i)]).$$

The proof is now completed by invoking the uniqueness of prime factorizations for one-variable polynomials with real coefficients. More precisely, note that we have exhibited factorizations of  $R'_\mu(x)$  and  $R'_\nu(x)$  into products of linear factors. These two monic polynomials are equal iff their linear factors (counting multiplicities) are the same, which holds iff the multisets

$$[\mu_i - (n-i) : 1 \leq i \leq n] \text{ and } [\nu_i - (n-i) : 1 \leq i \leq n]$$

are the same. Adding  $n$  to everything, this is equivalent to the multiset equality in the theorem statement.  $\square$

**12.11. Example.** The partitions  $(2, 2, 0, 0)$  and  $(3, 1, 0, 0)$  are rook-equivalent, because  $[3, 4, 3, 4] = [4, 3, 3, 4]$ . The partitions  $(4, 2, 1)$  and  $(5, 2)$  are not rook-equivalent, since  $[5, 4, 4, 4, 5, 6, 7] \neq [6, 4, 3, 4, 5, 6, 7]$ .

## 12.4 Parking Functions

This section illustrates the use of a probabilistic argument to enumerate a collection of combinatorial objects, namely the parking functions defined next.

**12.12. Definition: Parking Functions.** A *parking function of order  $n$*  is a function  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  such that

$$|\{x : f(x) \leq i\}| \geq i \quad \text{for } 1 \leq i \leq n.$$

**12.13. Example.** For  $n = 8$ , the function  $f$  defined in Figure 12.5 is a parking function, but the function  $g$  is not.

$x$	1	2	3	4	5	6	7	8
$f(x)$	2	6	3	2	6	2	2	1
$g(x)$	5	6	1	5	6	1	7	1

**FIGURE 12.5**

A parking function and a non-parking function.

The name “parking function” arises as follows. Consider a one-way street with  $n$  parking spaces numbered  $1, 2, \dots, n$ . Cars numbered  $1, 2, \dots, n$  arrive at the beginning of this street in numerical order. Each car wants to park in its own preferred spot on the street. We encode these parking preferences by a function  $h : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ , by letting  $h(x)$  be the parking spot preferred by car  $x$ . Given  $h$ , the cars park in the following way. For  $x = 1, 2, \dots, n$ , car  $x$  arrives and drives forward along the street to the spot  $h(x)$ . If that spot is empty, car  $x$  parks there. Otherwise, the car continues to drive forward on the one-way street and parks in the first available spot after  $h(x)$ , if any. The cars cannot return to the beginning of the street, so it is possible that not every car will be able to park.

For example, suppose the parking preferences are given by the parking function  $f$  defined in Figure 12.5. Car 1 arrives first and parks in spot 2. The next two cars arrive and park in spots 6 and 3, respectively. When car 4 arrives, spots 2 and 3 are full, so car 4 parks in spot 4. This process continues. At the end, every car has parked successfully, and the parking order is 8, 1, 3, 4, 6, 2, 5, 7. Now suppose the parking preferences are given by the non-parking function  $g$  defined in Figure 12.5. After the first six cars have arrived, the parking spots on the street are filled as follows:

$$3, 6, -, -, 1, 2, 4, 5.$$

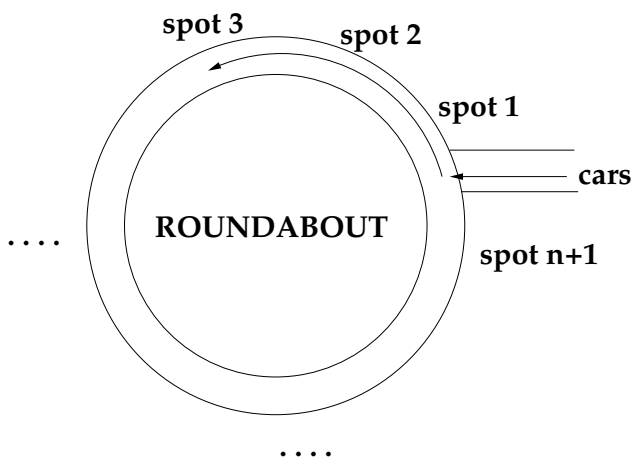
Car 7 arrives and drives to spot  $g(7) = 7$ . Since spots 7 and 8 are both full at this point, car 7 cannot park.

**12.14. Theorem.** A function  $h : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  is a parking function iff every car is able to park using the parking preferences determined by  $h$ .

*Proof.* We prove the contrapositive in each direction. Suppose first that  $h$  is not a parking function. Then there exists  $i \leq n$  such that  $|\{x : h(x) \leq i\}| < i$ . This means that fewer than  $i$  cars prefer to park in the first  $i$  spots. But then the first  $i$  spots cannot all be used, since a car never parks in a spot prior to the spot it prefers. Since there are  $n$  cars and  $n$  spots, the existence of an unused spot implies that not every car was able to park.

Conversely, assume not every car can park. Let  $i$  be the earliest spot that is not taken



**FIGURE 12.6**

Parking on a roundabout.

after every car has attempted to park. Then no car preferred spot  $i$ . Suppose  $i$  or more cars preferred the first  $i - 1$  spots. Not all of these cars can park in the first  $i - 1$  spots. But then one of these cars would have parked in spot  $i$ , a contradiction. We conclude that  $|\{x : h(x) \leq i\}| < i$ , so that  $h$  is not a parking function.  $\square$

**12.15. Theorem: Enumeration of Parking Functions.** There are  $(n + 1)^{n-1}$  parking functions of order  $n$ .

*Proof.* Fix  $n > 0$ . Define a *circular parking function* of order  $n$  to be any function  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n + 1\}$ . Let  $Z$  be the set of all such functions; we know that  $|Z| = (n + 1)^n$ . We interpret circular parking functions as follows. Imagine a roundabout (circular street) with  $n + 1$  parking spots numbered  $1, 2, \dots, n + 1$ . See Figure 12.6. As before,  $f$  encodes the parking preferences of  $n$  cars that wish to park on the roundabout. Thus, for  $1 \leq x \leq n$  and  $1 \leq y \leq n + 1$ ,  $y = f(x)$  iff car  $x$  prefers to park in spot  $y$ . Cars  $1, 2, \dots, n$  arrive at the roundabout in increasing order. Each car  $x$  enters just before spot 1, then drives around to spot  $f(x)$  and parks there if possible. If spot  $f(x)$  is full, car  $x$  keeps driving around the roundabout and parks in the first empty spot that it encounters.

No matter what  $f$  is, every car will succeed in parking in the circular situation. Moreover, since there are now  $n + 1$  spots and only  $n$  cars, there will always be one empty spot at the end. Suppose we randomly select a circular parking function. Because of the symmetry of the roundabout, each of the  $n + 1$  parking spaces is equally likely to be the empty one. (The fact that the entrance to the roundabout is at spot 1 is irrelevant here, since for parking purposes we may as well assume that car  $x$  enters the roundabout at its preferred spot  $f(x)$ .) Thus, the probability that spot  $k$  is empty is  $\frac{1}{n+1}$ , for  $1 \leq k \leq n + 1$ . On the other hand, spot  $n + 1$  will be the empty spot iff  $f$  is a parking function of order  $n$ . For, if spot  $n + 1$  is empty, then no car preferred spot  $n + 1$ , and no car passed spot  $n + 1$  during the parking process. Thus, the circular parking process on the roundabout coincides with the original parking process on the one-way street (and conversely). Since spot  $n + 1$  is empty with probability  $1/(n + 1)$  and the sample space  $Z$  has size  $(n + 1)^n$ , we conclude that the number of ordinary parking functions must be  $|Z|/(n + 1) = (n + 1)^{n-1}$ .  $\square$

**12.16. Remark.** Let  $A_{n,k}$  be the set of circular parking functions of order  $n$  with empty spot  $k$ . The preceding proof shows that  $|A_{n,k}| = (n + 1)^{n-1}$  for  $1 \leq k \leq n + 1$ . We

established this counting result by a probabilistic argument, using symmetry to deduce that  $P(A_{n,k}) = 1/(n+1)$  for all  $k$ . This symmetry property is intuitively evident, but it can also be proved rigorously as follows. Suppose  $f \in A_{n,k_1}$  and  $k_2$  are given. Let  $\phi(f)$  be the function  $i \mapsto f(i) + k_2 - k_1 \bmod (n+1)$  for  $1 \leq i \leq n$ , taking the remainder to lie in  $\{1, 2, \dots, n+1\}$ . One may check that  $\phi$  defines a bijection from  $A_{n,k_1}$  onto  $A_{n,k_2}$ . These bijections prove that all the sets  $A_{n,k}$  (for  $1 \leq k \leq n+1$ ) have the same cardinality.

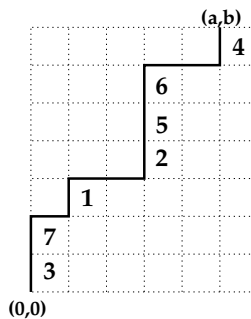
**12.17. Remark.** One of the early motivations for studying parking functions was their connection to hashing protocols. In computing applications, one often stores information in a data structure called a *hash table*. We consider a simplified model where  $n$  items are to be stored in a linear array of  $n$  cells. A *hash function*  $h : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  is used to determine where each item will be stored. We store item  $i$  in position  $h(i)$ , *unless* that position has already been taken by a previous item — this circumstance is called a *collision*. We handle collisions via the following *collision resolution policy*: if  $h(i)$  is full, we store item  $i$  in the earliest position after position  $i$  that is not yet full (if any). If there is no such position, the collision resolution fails (we do not allow “wraparound”). This scenario is exactly like that of the cars parking on a one-way street according to the preferences encoded by  $h$ . Thus, we will be able to store all  $n$  items in the hash table iff  $h$  is a parking function.

## 12.5 Parking Functions and Trees

We can use parking functions (§12.4) to give a bijective proof of Cayley’s formula 3.72 for the number of  $n$ -vertex trees. The proof involves labeled lattice paths, which we now define.

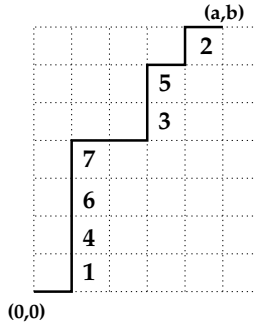
**12.18. Definition: Labeled Lattice Paths.** A *labeled lattice path* consists of a lattice path  $\pi$  from  $(0, 0)$  to  $(a, b)$ , together with a labeling of the  $b$  north steps of  $\pi$  with labels  $1, 2, \dots, b$  (each used exactly once) such that the labels for the north steps in a given column increase from bottom to top.

We can illustrate a labeled lattice path by drawing  $\pi$  inside an  $(a+1) \times b$  grid of unit squares and placing the label of each north step in the unit square to the right of that north step. For example, Figure 12.7 displays a labeled lattice path ending at  $(5, 7)$ .



**FIGURE 12.7**

A labeled lattice path.

**FIGURE 12.8**

Converting a function to a labeled path.

**12.19. Theorem: Enumeration of Labeled Paths.** There are  $(a+1)^b$  labeled lattice paths from  $(0,0)$  to  $(a,b)$ .

*Proof.* It suffices to construct a bijection between the set of labeled lattice paths ending at  $(a,b)$  and the set of all functions  $f : \{1, 2, \dots, b\} \rightarrow \{1, 2, \dots, a+1\}$ . Given a labeled lattice path  $P$ , define the associated function by setting  $f(i) = j$  for all labels  $i$  in column  $j$  of  $P$ .

The inverse map acts as follows. Given a function  $f : \{1, 2, \dots, b\} \rightarrow \{1, 2, \dots, a+1\}$ , let  $S_i = \{x : f(x) = i\}$  and  $s_i = |S_i|$  for  $1 \leq i \leq a+1$ . The labeled path associated to  $f$  is the lattice path  $N^{s_1} E N^{s_2} E \dots N^{s_{a+1}}$  where the  $i$ th string of consecutive north steps is labeled by the elements of  $S_i$  in increasing order.  $\square$

**12.20. Example.** The function associated to the labeled path  $P$  in Figure 12.7 is given by

$$1 \mapsto 2, 2 \mapsto 4, 3 \mapsto 1, 4 \mapsto 6, 5 \mapsto 4, 6 \mapsto 4, 7 \mapsto 1.$$

Going the other way, the function  $f : \{1, 2, \dots, 7\} \rightarrow \{1, 2, \dots, 6\}$  defined by

$$f(1) = 2, f(2) = 5, f(3) = 4, f(4) = 2, f(5) = 4, f(6) = 2, f(7) = 2,$$

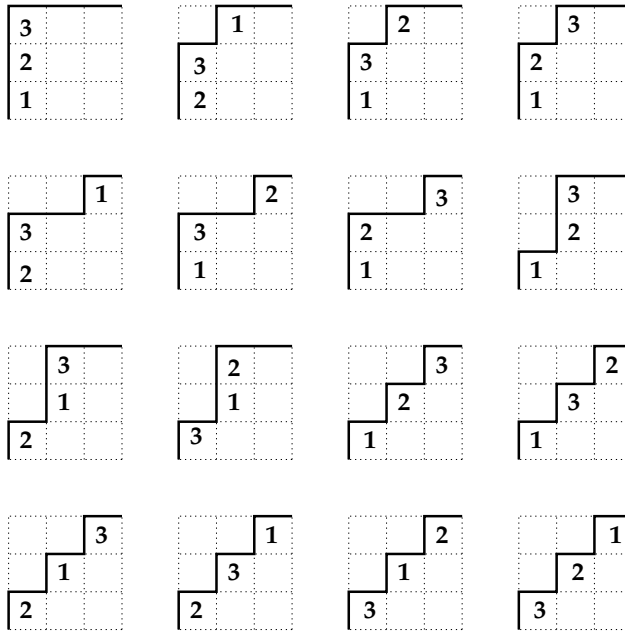
is mapped to the labeled lattice path shown in Figure 12.8.

A *labeled Dyck path of order  $n$*  is a Dyck path ending at  $(n,n)$  that is labeled according to the rules in 12.18. For example, Figure 12.9 displays the sixteen labeled Dyck paths of order 3.

**12.21. Theorem: Enumeration of Labeled Dyck Paths.** There are  $(n+1)^{n-1}$  labeled Dyck paths of order  $n$ .

*Proof.* Using the bijection in 12.19, we can regard labeled lattice paths from  $(0,0)$  to  $(n,n)$  as functions  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n+1\}$ . We first show that non-Dyck labeled paths correspond to non-parking functions under this bijection. A labeled path  $P$  is *not* a Dyck path iff some east step of  $P$  goes from  $(i-1, j)$  to  $(i, j)$  for some  $i > j$ . This condition holds for  $P$  iff the function  $f$  associated to  $P$  satisfies  $|\{x : f(x) \leq i\}| = j$  for some  $i > j$ . In turn, this condition on  $f$  is equivalent to the existence of  $i$  such that  $|\{x : f(x) \leq i\}| < i$ . But this means that  $f$  is *not* a parking function (see 12.12). It now follows that labeled Dyck paths are in bijective correspondence with parking functions. So the result follows from 12.15.  $\square$

**12.22. Cayley's Theorem via Parking Functions.** There are  $(n+1)^{n-1}$  trees with vertex set  $\{0, 1, 2, \dots, n\}$ .

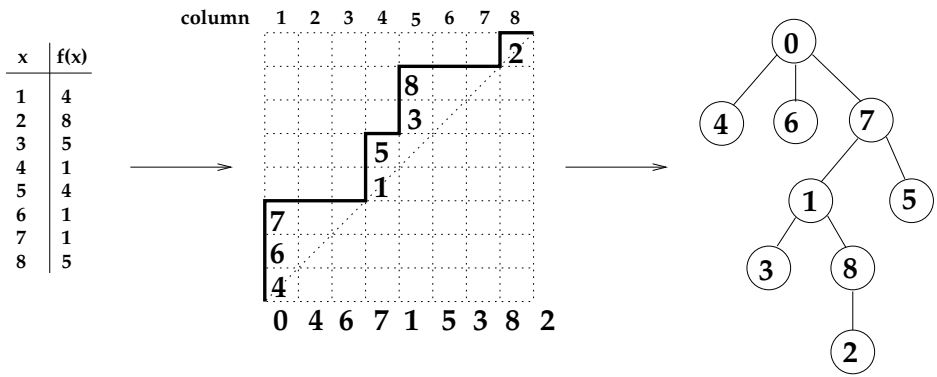


**FIGURE 12.9**  
Labeled Dyck paths.

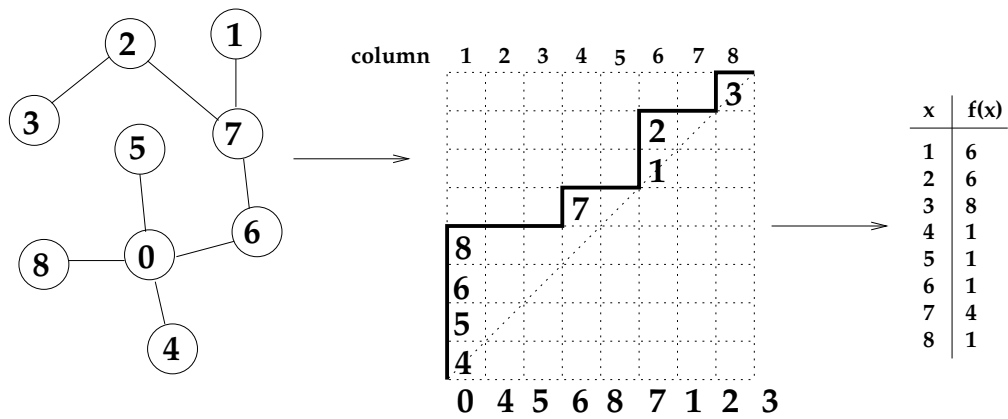
*Proof.* In light of the previous result, it suffices to define bijections between the set  $B$  of labeled Dyck paths of order  $n$  and the set  $C$  of all trees with vertex set  $\{0, 1, 2, \dots, n\}$ . To define  $f : B \rightarrow C$ , let  $\pi$  be a labeled Dyck path of order  $n$ . Let  $(a_1, a_2, \dots, a_n)$  be the sequence of labels in the diagram of  $\pi$ , reading from the bottom row to the top row, and set  $a_0 = 0$ . Define a graph  $T = f(\pi)$  as follows. For  $0 \leq j \leq n$ , there is an edge in  $T$  from vertex  $a_j$  to each vertex whose label appears in column  $j + 1$  of the diagram of  $\pi$ . These are all the edges of  $T$ . Using the fact that  $\pi$  is a labeled Dyck path, one proves by induction on  $j$  that every  $a_j$  is either 0 or appears to the left of column  $j + 1$  in the diagram, so that every vertex in column  $j + 1$  of  $\pi$  is reachable from vertex 0 in  $T$ . Thus,  $T = f(\pi)$  is a connected graph with  $n$  edges and  $n + 1$  vertices, so  $T$  is a tree by 3.71.

**12.23. Example.** Figure 12.10 shows a parking function  $f$ , the labeled Dyck path  $\pi$  corresponding to  $f$ , and the tree  $T = f(\pi)$ . We can use the figure to compute the edges of  $T$  by writing  $a_j$  underneath column  $j + 1$ , for  $0 \leq j \leq n$ . If we regard zero as the ancestor of all other vertices, then the labels in column  $j + 1$  are the children of vertex  $a_j$ .

Continuing the proof, we define the inverse map  $f' : C \rightarrow B$ . Let  $T \in C$  be a tree with vertex set  $\{0, 1, 2, \dots, n\}$ . We generate the diagram for  $f'(T)$  by inserting labels into an  $n \times n$  grid from bottom to top. Denote these labels by  $(a_1, \dots, a_n)$ , and set  $a_0 = 0$ . The labels  $a_1, a_2, \dots$  in column 1 are the vertices of  $T$  adjacent to vertex  $a_0 = 0$  (written in increasing order from bottom to top). The labels in the second column are the neighbors of  $a_1$  other than vertex 0. The labels in the third column are the neighbors of  $a_2$  not in the set  $\{a_0, a_1\}$ . In general, the labels in column  $j + 1$  are the neighbors of  $a_j$  not in the set  $\{a_0, a_1, \dots, a_{j-1}\}$ . Observe that we do not know the full sequence  $(a_1, \dots, a_n)$  in advance, but we reconstruct this sequence as we go along. We will show momentarily that  $a_j$  is always known by the time we reach column  $j + 1$ .



**FIGURE 12.10**  
Mapping parking functions to labeled Dyck paths to trees.



**FIGURE 12.11**  
Mapping trees to labeled Dyck paths to parking functions.

**12.24. Example.** Figure 12.11 shows a tree  $T$ , the labeled Dyck path  $\pi = f'(T)$ , and the parking function associated to  $\pi$ .

Let us check that  $f'$  is well defined. We break up the computation of  $f'(T)$  into stages, where stage  $j$  consists of choosing the increasing sequence of labels  $a_i < \cdots < a_k$  that occur in column  $j$ . We claim that at each stage  $j$  with  $1 \leq j \leq n+1$ ,  $a_{j-1}$  has already been computed, so that the labels entered in column  $j$  occur in rows  $j$  or higher. This will show that the algorithm for computing  $f'$  is well defined and produces a labeled Dyck path. We proceed by induction on  $j$ . The claim holds for  $j = 1$ , since  $a_0 = 0$  by definition. Assume that  $1 < j \leq n+1$  and that the claim holds for all  $j' < j$ . To get a contradiction, assume that  $a_{j-1}$  is not known when we reach column  $j$ . Since the claim holds for  $j-1$ , we must have already recovered the labels in the set  $W = \{a_0 = 0, a_1, \dots, a_{j-2}\}$ , which are precisely the labels that occur in the first  $j-1$  columns. Let  $z$  be a vertex of  $T$  not in  $W$ . Since  $T$  is a tree, there is a path from 0 to  $z$  in  $T$ . Let  $y$  be the earliest vertex on this path not in  $W$ , and let  $x$  be the vertex just before  $y$  on the path. By choice of  $y$ , we have  $x \in W$ , so that  $x = 0$  or  $x = a_k$  for some  $k \leq j-2$ . But if  $x = 0$ , then  $y$  occurs in column 1 and hence  $y \in W$ . And if  $x = a_k$ , then the algorithm for  $f'$  would have placed  $y$  in column  $k+1 \leq j-1$ , and again  $y \in W$ . These contradictions show that the claim holds for  $j$ . It is now routine to check that  $f'$  is the two-sided inverse of  $f$ .  $\square$

## 12.6 Möbius Inversion and Field Theory

This section gives two applications of the material in §4.7 to field theory. We show that every finite subgroup of the multiplicative group of any field must be cyclic; and we count the number of irreducible polynomials of a given degree with coefficients in a given finite field. The starting point for proving the first result is the relation  $n = \sum_{d|n} \phi(d)$ , proved in 4.34. We begin by giving a combinatorial interpretation of this identity in terms of the orders of elements in a cyclic group of size  $n$ .

**12.25. Theorem: Order of Elements in a Cyclic Group.** Suppose  $G$  is a cyclic group of size  $d < \infty$ , written multiplicatively. If  $x \in G$  generates  $G$  and  $c \geq 1$ , then  $x^c$  generates a cyclic subgroup of  $G$  of order  $d/\gcd(c, d) = \text{lcm}(c, d)/c$ .

*Proof.* Since  $\langle x^c \rangle \subseteq G$ , the order of  $x^c$  must be finite. Let  $k$  be the order of  $x^c$ . We have seen in 9.79 that  $k$  is the smallest positive integer such that  $x^{ck} = 1_G$ , and that the  $k$  elements  $x^c, x^{2c}, \dots, x^{kc}$  are distinct and constitute the cyclic subgroup of  $G$  generated by  $x^c$ . Since  $x$  has order  $d$ , we know from 9.79 that  $x^m = 1$  iff  $d|m$ . It follows from this and the definition of  $k$  that  $kc$  is the least positive multiple of  $c$  that is also a multiple of  $d$ . In other words,  $kc = \text{lcm}(c, d)$ . It follows that the order of  $x^c$  is  $k = \text{lcm}(c, d)/c$ . Since  $cd = \text{lcm}(c, d)\gcd(c, d)$ , we also have  $k = d/\gcd(c, d)$ .  $\square$

**12.26. Theorem: Counting Generators in a Cyclic Group.** If  $G$  is a cyclic group of size  $d < \infty$ , then there are exactly  $\phi(d)$  elements in  $G$  that generate  $G$ .

*Proof.* Let  $x$  be a fixed generator of  $G$ . By 9.79, the  $d$  distinct elements of  $G$  are  $x^1, x^2, \dots, x^d = 1_G$ . By 12.25, the element  $x^c$  generates all of  $G$  iff  $\gcd(c, d) = 1$ . By the definition of  $\phi$  (see 4.19), the number of such integers  $c$  between 1 and  $d$  is precisely  $\phi(d)$ .  $\square$

**12.27. Theorem: Subgroup Structure of Cyclic Groups.** Let  $G$  be a cyclic group

of size  $n < \infty$ . For each  $d$  dividing  $n$ , there exists exactly one subgroup of  $G$  of size  $d$ , and this subgroup is cyclic.

*Proof.* We only sketch the proof, which uses some results about group homomorphisms that were stated as exercises in Chapter 9. We know from 9.59 that every subgroup of the cyclic group  $\mathbb{Z}$  has the form  $k\mathbb{Z}$  for some unique  $k \geq 0$ , and is therefore cyclic. Next, any finite cyclic group  $G$  can be viewed as the quotient group  $\mathbb{Z}/n\mathbb{Z}$  for some  $n \geq 1$ . This follows by applying the fundamental homomorphism theorem 9.207 to the map from  $\mathbb{Z}$  to  $G$  sending 1 to a generator of  $G$ . By the correspondence theorem 9.211, each subgroup  $H$  of  $G$  has the form  $H = m\mathbb{Z}/n\mathbb{Z}$  for some subgroup  $m\mathbb{Z}$  of  $\mathbb{Z}$  containing  $n\mathbb{Z}$ . Now,  $m\mathbb{Z}$  contains  $n\mathbb{Z}$  iff  $m|n$ , and in this case  $|m\mathbb{Z}/n\mathbb{Z}| = n/m$ . It follows that there is a bijection between the positive divisors of  $n$  and the subgroups of  $G$ . Each such subgroup is the homomorphic image of a cyclic group  $m\mathbb{Z}$ , so each subgroup of  $G$  is cyclic.  $\square$

Suppose  $G$  is cyclic of size  $n$ . For each  $d|n$ , let  $G_d$  be the unique (cyclic) subgroup of  $G$  of size  $d$ . On one hand, each element  $y$  of  $G$  generates exactly one of the subgroups  $G_d$  (namely,  $y$  generates the group  $G_d$  such that  $d$  is the order of  $y$ ). On the other hand, we have shown that  $G_d$  has exactly  $\phi(d)$  generators. Invoking the sum rule, we obtain a new proof of the fact that

$$n = \sum_{d|n} \phi(d).$$

**12.28. Theorem: Detecting Cyclic Groups.** If  $G$  is a group of size  $n$  such that for each  $d$  dividing  $n$ ,  $G$  has at most one subgroup of size  $d$ , then  $G$  is cyclic.

*Proof.* For each  $d$  dividing  $n$ , let  $T_d$  be the set of elements in  $G$  of order  $d$ .  $G$  is the disjoint union of the sets  $T_d$  by 9.119. Consider a fixed choice of  $d$  such that  $T_d$  is nonempty. Then  $G$  has an element of order  $d$ , hence has a cyclic subgroup of size  $d$ . By assumption, this is the only subgroup of  $G$  of size  $d$ , and we know this subgroup has  $\phi(d)$  generators. Therefore,  $|T_d| = \phi(d)$  whenever  $|T_d| \neq 0$ . We conclude that

$$n = |G| = \sum_{d|n} |T_d| \leq \sum_{d|n} \phi(d) = n.$$

Since the extreme ends of this calculation both equal  $n$ , the middle inequality here must in fact be an equality. This is only possible if every  $T_d$  is nonempty. In particular,  $T_n$  is nonempty. Therefore,  $G$  is cyclic, since it is generated by each of the elements in  $T_n$ .  $\square$

**12.29. Theorem: Multiplicative Subgroups of Fields.** Let  $F$  be any field, possibly infinite. If  $G$  is a finite subgroup of the multiplicative group of  $F$ , then  $G$  is cyclic.

*Proof.* Suppose  $G$  is a subgroup of  $F^*$  (the multiplicative group of nonzero elements in  $F$ ) such that  $|G| = n < \infty$ . By 12.28, it suffices to show that  $G$  has at most one subgroup of size  $d$ , for each  $d|n$ . If not, let  $H$  and  $K$  be two distinct subgroups of  $G$  of size  $d$ . Then  $H \cup K$  is a set with at least  $d + 1$  elements; and for each  $z \in H \cup K$ , it follows from 9.119 that  $z$  is a root of the polynomial  $x^d - 1$  in  $F$ . But any polynomial of degree  $d$  over  $F$  has at most  $d$  distinct roots in the field  $F$  (see 2.157). This contradiction completes the proof.  $\square$

Our next goal is to count irreducible polynomials of a given degree over a finite field. We shall assume a number of results from field theory, whose proofs may be found in Chapter V of the algebra text by Hungerford [70]. Let  $F$  be a finite field with  $q$  elements. It is known that  $q$  must be a prime power, say  $q = p^e$ , and  $F$  is uniquely determined (up to isomorphism) by its cardinality  $q$ . Every finite field  $F$  with  $q = p^e$  elements is a splitting field for the polynomial  $x^q - x$  over  $\mathbb{Z}/p\mathbb{Z}$ .

**12.30. Theorem: Enumeration of Irreducible Polynomials.** Let  $F$  be a field with  $q = p^e$  elements. For each  $n \geq 1$ , let  $I(n, q)$  be the number of monic irreducible polynomials of degree  $n$  in the polynomial ring  $F[x]$ . Then

$$q^n = \sum_{d|n} dI(d, q)$$

and hence

$$I(n, q) = \frac{1}{n} \sum_{d|n} q^d \mu(n/d).$$

*Proof.* The strategy of the proof is to classify the elements in a finite field  $K$  of size  $q^n$  based on their minimal polynomials. From field theory, we know that each element  $u \in K$  is the root of a uniquely determined monic, irreducible polynomial in  $F[x]$  (called the *minimal polynomial of  $u$  over  $F$* ). The degree  $d$  of this minimal polynomial is  $d = [F(u) : F]$ , where for any field extension  $E \subseteq H$ ,  $[H : E]$  denotes the dimension of  $H$  viewed as a vector space over  $E$ . It is known that  $n = [K : F] = [K : F(u)] \cdot [F(u) : F]$ , so that  $d|n$ . Conversely, given any divisor  $d$  of  $n$ , we claim that every irreducible polynomial of degree  $d$  in  $F[x]$  has  $d$  distinct roots in  $K$ . Sketch of proof: Suppose  $g$  is such a polynomial and  $z \neq 0$  is a root of  $g$  in a splitting field of  $g$  over  $K$ . Since  $z$  lies in  $F(z)$ , which is a field with  $q^d$  elements, it follows from 9.119 (applied to the multiplicative group  $F(z)^*$ ) that  $z^{q^d-1} = 1$ . One checks that  $q^d - 1$  divides  $q^n - 1$  (since  $d|n$ ), so that  $z^{q^n-1} = 1$ , and hence  $z$  is a root of  $x^{q^n} - x$ . It follows that every root  $z$  of  $g$  actually lies in  $K$  (which is a splitting field for  $x^{q^n} - x$ ). Furthermore, since  $z$  is a root of  $x^{q^n} - x$ , it follows that the minimal polynomial for  $z$  over  $F$  (namely  $g$ ) divides  $x^{q^n} - x$  in  $F[x]$ . We conclude that  $g$  divides  $x^{q^n} - x$  in  $K[x]$  also. The polynomial  $x^{q^n} - x$  is known to split into a product of  $q^n$  *distinct* linear factors over  $K$ ; in fact,  $x^{q^n} - x = \prod_{x_0 \in K} (x - x_0)$ . By unique factorization in the polynomial ring  $K[x]$ ,  $g$  must also be a product of  $d$  *distinct* linear factors. This completes the proof of the claim.

We can now write  $K$  as the disjoint union of sets  $R_g$  indexed by all irreducible polynomials in  $F[x]$  whose degrees divide  $n$ , where  $R_g$  consists of the  $\deg(g)$  distinct roots of  $g$  in  $K$ . Invoking the sum rule and grouping together terms indexed by polynomials of degree  $d$  dividing  $n$ , we obtain

$$q^n = |K| = \sum_{\substack{\text{irreducible } g \\ \deg(g)|n}} |R_g| = \sum_{\substack{\text{irreducible } g \\ \deg(g)|n}} \deg(g) = \sum_{d|n} dI(d, q).$$

We can now apply the Möbius inversion formula 4.30 to the functions  $f(n) = q^n$  and  $g(n) = nI(n, q)$  to obtain

$$nI(n, q) = \sum_{d|n} q^d \mu(n/d). \quad \square$$

## 12.7 Quantum Binomial Coefficients and Subspaces

Recall (§6.7) that the *quantum binomial coefficients* are the polynomials in  $\mathbb{N}[x]$  defined by the formula

$$\begin{bmatrix} n \\ k \end{bmatrix}_x = \frac{[n]!_x}{[k]!_x [n-k]!_x} = \frac{\prod_{i=1}^n (x^i - 1)}{\prod_{i=1}^k (x^i - 1) \prod_{i=1}^{n-k} (x^i - 1)}.$$

We gave a number of combinatorial interpretations of these polynomials in §6.7. In this section, we discuss a linear-algebraic interpretation of the integers  $\begin{bmatrix} n \\ k \end{bmatrix}_q$ , where  $q$  is a prime



power. To read this section, the reader should have some previous experience with fields and vector spaces. We begin by using bases to determine the possible sizes of vector spaces over finite fields.

**12.31. Theorem: Size of Vector Spaces over Finite Fields.** Suppose  $V$  is a  $d$ -dimensional vector space over a finite field  $F$  with  $q$  elements. Then  $|V| = q^d$ .

*Proof.* Let  $(v_1, \dots, v_d)$  be an ordered basis for  $V$ . By definition of a basis, for each  $v \in V$ , there exists exactly one  $d$ -tuple of scalars  $(c_1, \dots, c_d) \in F^d$  such that  $v = c_1 v_1 + c_2 v_2 + \dots + c_d v_d$ . In other words, there is a bijection  $v \mapsto (c_1, \dots, c_d)$  from  $V$  to  $F^d$ . Because  $|F| = q$ , the product rule gives  $|V| = |F^d| = |F|^d = q^d$ .  $\square$

**12.32. Theorem: Size of Finite Fields.** If  $K$  is a finite field, then  $|K| = p^e$  for some prime  $p$  and some  $e \geq 1$ .

*Proof.* Given  $K$ , let  $F$  be the cyclic subgroup of the additive group  $(K, +)$  generated by  $1_K$ . The size of  $F$  is some finite number  $p$  (since  $K$  is finite), and  $p > 1$  since  $1_K \neq 0_K$ . We know that  $p$  is the smallest positive integer such that  $p1_K = 0_K$ . One checks (using the distributive laws) that  $F$  is a subring of  $K$ , not just a subgroup. If  $p$  were not prime, say  $p = ab$  with  $1 < a, b < p$ , then  $(a1_K) \cdot (b1_K) = ab1_K = p1_K = 0_K$ , and yet  $a1_K, b1_K \neq 0$ . This contradicts the fact that fields have no zero divisors. Thus,  $p$  must be prime. It now follows that  $F$  is a field isomorphic to the field of integers modulo  $p$ .  $K$  can be regarded as a vector space over its subfield  $F$ , by defining scalar multiplication  $F \times K \rightarrow K$  to be the restriction of the multiplication  $K \times K \rightarrow K$  in the field. Since  $K$  is finite, it must be a finite-dimensional vector space over  $F$ . Thus the desired result follows from 12.31.  $\square$

**12.33. Remark.** One can show that, for every prime power  $p^e$ , there exists a finite field of size  $p^e$ , which is unique up to isomorphism. The existence proof is sketched in 12.126.

We now give the promised linear-algebraic interpretation of quantum binomial coefficients.

**12.34. Theorem.** Let  $K$  be a finite field with  $q$  elements. For all integers  $n \geq 0$  and  $0 \leq k \leq n$ ,  $\begin{bmatrix} n \\ k \end{bmatrix}_q$  is the number of  $k$ -dimensional subspaces of any  $n$ -dimensional vector space  $V$  over  $K$ .

*Proof.* Let  $f(n, k, q)$  be the number of  $k$ -dimensional subspaces of  $V$ . (One can check that this number depends only on  $k$ ,  $q$ , and  $n = \dim(V)$ .) Recall from 12.31 that  $|V| = q^n$  and each  $d$ -dimensional subspace of  $V$  has size  $q^d$ . By rearranging factors in the defining formula for  $\begin{bmatrix} n \\ k \end{bmatrix}_q$ , we see that  $\begin{bmatrix} n \\ k \end{bmatrix}_q = f(n, k, q)$  holds iff

$$f(n, k, q)(q^k - 1)(q^{k-1} - 1) \cdots (q^1 - 1) = (q^n - 1)(q^{n-1} - 1) \cdots (q^{n-k+1} - 1).$$

We establish this equality by the following counting argument. Let  $S$  be the set of all ordered lists  $(v_1, \dots, v_k)$  of  $k$  linearly independent vectors in  $V$ . Here is one way to build such a list. First, choose a nonzero vector  $v_1 \in V$  in any of  $q^n - 1$  ways. This vector spans a one-dimensional subspace  $W_1$  of  $V$  of size  $q = q^1$ . Second, choose a vector  $v_2 \in V \sim W_1$  in any of  $q^n - q$  ways. The list  $(v_1, v_2)$  must be linearly independent since  $v_2$  is not in the space  $W_1$  spanned by  $v_1$ . Vectors  $v_1$  and  $v_2$  span a two-dimensional subspace  $W_2$  of  $V$  of size  $q^2$ . Third, choose  $v_3 \in V \sim W_2$  in  $q^n - q^2$  ways. Continue similarly. When choosing  $v_i$ , we have already found  $i - 1$  linearly independent vectors  $v_1, \dots, v_{i-1}$  that span a subspace of  $V$  of size  $q^{i-1}$ . Consequently,  $(v_1, \dots, v_i)$  will be linearly independent iff we choose  $v_i \in V \sim W_i$ , which is a set of size  $q^n - q^{i-1}$ . By the product rule, we conclude that

$$|S| = \prod_{i=1}^k (q^n - q^{i-1}) = \prod_{i=1}^k q^{i-1} (q^{n+1-i} - 1) = q^{k(k-1)/2} (q^n - 1)(q^{n-1} - 1) \cdots (q^{n-k+1} - 1).$$

Now let us count  $S$  in a different way. Observe that the vectors in each list  $(v_1, \dots, v_k) \in S$  span some  $k$ -dimensional subspace of  $V$ . So we can begin by choosing such a subspace  $W$  in any of  $f(n, k, q)$  ways. Next we choose  $v_1, \dots, v_k \in W$  one at a time, following the same process used in the first part of the proof. We can choose  $v_1$  in  $|W| - 1 = q^k - 1$  ways, then  $v_2$  in  $q^k - q$  ways, and so on. By the product rule,

$$|S| = f(n, k, q) \prod_{i=1}^k (q^k - q^{i-1}) = f(n, k, q) q^{k(k-1)/2} (q^k - 1)(q^{k-1} - 1) \cdots (q^1 - 1).$$

Equating the two formulas for  $|S|$  and cancelling  $q^{k(k-1)/2}$  gives the desired result.  $\square$

In 6.36, we saw that

$$\begin{bmatrix} n \\ k \end{bmatrix}_x = \sum_{\mu \in P(k, n-k)} x^{|\mu|}$$

where  $P(k, n-k)$  is the set of all integer partitions  $\mu$  that fit in a  $k \times (n-k)$  rectangle. In the rest of this section, we shall give a second proof of 12.34 by showing that

$$f(n, k, q) = \sum_{\mu \in P(k, n-k)} q^{|\mu|}.$$

This proof is longer than the one just given, but it reveals a close connection between enumeration of subspaces on one hand, and enumeration of partitions in a box (or, equivalently, lattice paths) on the other hand.

For convenience, we shall work with the vector space  $V = K^n$  whose elements are  $n$ -tuples of elements of  $K$ . We regard elements of  $V$  as row vectors of length  $n$ . The key linear-algebraic fact we need is that every  $k$ -dimensional subspace of  $V = K^n$  has a unique “reduced row-echelon form basis.”

**12.35. Definition: Reduced Row-Echelon Form.** Let  $A$  be a  $k \times n$  matrix with entries in  $K$ . Let  $A_1, \dots, A_k \in K^n$  be the  $k$  rows of  $A$ . We say  $A$  is a *reduced row-echelon form* (RREF) matrix iff the following conditions hold: (i)  $A_i \neq 0$  for all  $i$ , and the leftmost nonzero entry of  $A_i$  is  $1_K$  (call these entries *leading ones*); (ii) if the leading one of  $A_i$  occurs in column  $j(i)$ , then  $j(1) < j(2) < \cdots < j(k)$ ; (iii) every leading one is the only nonzero entry in its column. An ordered basis  $B = (v_1, \dots, v_k)$  for a  $k$ -dimensional subspace of  $K^n$  is called a *RREF basis* iff the matrix whose rows are  $v_1, \dots, v_k$  is a RREF matrix.

**12.36. Theorem: RREF Bases.** Let  $K$  be any field. Every  $k$ -dimensional subspace of  $K^n$  has a unique RREF basis. Conversely, the rows of every  $k \times n$  RREF matrix comprise an ordered basis for a  $k$ -dimensional subspace of  $K^n$ . Consequently, there is a bijection between the set of such subspaces and the set of  $k \times n$  RREF matrices with entries in  $K$ .

*Proof.* We sketch the proof, trusting the reader’s ability to supply the remaining linear algebra details. *Step 1:* We use row-reduction to show that any given  $k$ -dimensional subspace  $W$  of  $K^n$  has *at least one* RREF basis. Start with any ordered basis  $v_1, \dots, v_k$  of  $W$ , and let  $A$  be the matrix with rows  $v_1, \dots, v_k$ . There are three “elementary row operations” we can use to simplify  $A$ : interchange two rows; multiply one row by a nonzero scalar; add any scalar multiple of one row to a different row. A routine verification shows that performing any one of these operations has no effect on the subspace spanned by the rows of  $A$ . Therefore, we can create new ordered bases for  $W$  by performing sequences of row operations on  $A$ . Using the well-known Gaussian elimination algorithm (“row reduction”), we can bring the matrix  $A$  into reduced row-echelon form. The rows of the new matrix give the desired RREF basis of  $W$ .

*Step 2:* We show that a given subspace  $W$  has *at most one* RREF basis. Use induction on  $k$ , the base case  $k = 0$  being immediate. For the induction step, assume  $n \geq 1$  and  $k \geq 1$  are fixed, and the uniqueness result is known for smaller values of  $k$ . Let  $A$  and  $B$  be two RREF matrices whose rows form bases of  $W$ ; we must prove  $A = B$ . Let  $j(1) < j(2) < \cdots < j(k)$  be the positions of the leading ones in  $A$ , and let  $r(1) < \cdots < r(k)$  be the positions of the leading ones in  $B$ . If  $j(1) < r(1)$ , then the first row of  $A$  (which is a vector in  $W$ ) has a 1 in position  $j(1)$ . This vector cannot possibly be a linear combination of the rows of  $B$ , all of whose nonzero entries occur in columns after  $j(1)$ . Thus,  $j(1) < r(1)$  is impossible. A similar argument rules out  $r(1) < j(1)$ , so we must have  $j(1) = r(1)$ . Let  $W'$  be the subspace of  $W$  consisting of vectors with zeroes in positions  $1, 2, \dots, j(1)$ . Consideration of leading ones shows that rows 2 through  $k$  of  $A$  must form a basis for  $W'$ , and rows 2 through  $k$  of  $B$  also form a basis for  $W'$ . Since  $\dim(W') = k - 1$ , the induction hypothesis implies that rows 2 through  $k$  of  $A$  equal the corresponding rows of  $B$ . In particular, we now know that  $r(i) = j(i)$  for  $1 \leq i \leq k$ . To finish, we must still check that row 1 of  $A$  equals row 1 of  $B$ . Let the rows of  $B$  be  $w_1, \dots, w_k$ , and write  $v_1$  for the first row of  $A$ . Since  $v_1 \in W$ , we have  $v_1 = a_1 w_1 + \cdots + a_k w_k$  for suitable scalars  $a_k$ . Consideration of column  $j(1)$  shows that  $a_1 = 1$ . On the other hand, if  $a_i \neq 0$  for some  $i > 1$ , then  $a_1 w_1 + \cdots + a_k w_k$  would have a nonzero entry in position  $j(i)$ , whereas  $v_1$  has a zero entry in this position (since the leading ones occur in the same columns in  $A$  and  $B$ ). This is a contradiction, so  $a_2 = \cdots = a_k = 0$ . Thus  $v_1 = w_1$ , as desired, and we have now proved that  $A = B$ .

*Step 3:* We show that the  $k$  rows  $v_1, \dots, v_k$  of a given RREF matrix form an ordered basis for some  $k$ -dimensional subspace of  $K^n$ . It suffices to show that the rows in question are linearly independent vectors. Suppose  $c_1 v_1 + \cdots + c_k v_k = 0$ , where  $c_i \in K$ . Recall that the leading one in position  $(i, j(i))$  is the only nonzero entry in its column. Therefore, taking the  $j(i)$ th component of the preceding equation, we get  $c_i = 0$  for  $1 \leq i \leq k$ . □

Because of the preceding theorem, the problem of counting  $k$ -dimensional subspaces of  $K^n$  (where  $|K| = q$ ) reduces to the problem of counting  $k \times n$  RREF matrices with entries in  $K$ . Our second proof of 12.34 will therefore be complete once we prove the following result.

**12.37. Theorem: Enumeration of RREF Matrices.** Let  $K$  be a finite field with  $q$  elements. The number of  $k \times n$  RREF matrices with entries in  $K$  is

$$\sum_{\mu \in P(k, n-k)} q^{|\mu|} = \begin{bmatrix} n \\ k \end{bmatrix}_q.$$

*Proof.* Let us classify the  $k \times n$  RREF matrices based on the columns  $j(1) < j(2) < \cdots < j(k)$  where the leading ones occur. To build a RREF matrix with the leading ones in these positions, we must put zeroes in all matrix positions  $(i, p)$  such that  $p < j(i)$ ; we must also put zeroes in all matrix positions  $(r, j(i))$  such that  $r < i$ . However, in all the other positions to the right of the leading ones, there is no restriction on the elements that occur except that they must come from the field  $K$  of size  $q$ . How many such “free positions” are there? The first row contains  $n - j(1)$  entries after the leading one, but  $k - 1$  of these entries are in columns above other leading ones. So there are  $\mu_1 = n - j(1) - (k - 1)$  free positions in this row. The next row contains  $n - j(2)$  entries after the leading one, but  $k - 2$  of these occur in columns above other leading ones. So there are  $\mu_2 = n - j(2) - (k - 2)$  free positions in row 2. Similarly, there are  $\mu_i = n - j(i) - (k - i) = n - k + i - j(i)$  free positions in row  $i$  for  $1 \leq i \leq k$ . The condition  $1 \leq j(1) < j(2) < \cdots < j(k) \leq n$  is logically equivalent to  $0 \leq j(1) - 1 \leq j(2) - 2 \leq \cdots \leq j(k) - k \leq n - k$ , which is in turn equivalent to  $n - k \geq \mu_1 \geq \mu_2 \geq \cdots \geq \mu_k \geq 0$ . Thus there is a bijection between the set of valid

positions  $j(1) < j(2) < \dots < j(k)$  for the leading ones, and the set of integer partitions  $\mu = (\mu_1, \dots, \mu_k)$  that fit in a  $k \times (n - k)$  box, given by  $\mu_i = n - k + i - j(i)$ . Furthermore,  $|\mu| = \mu_1 + \dots + \mu_k$  is the total number of free positions in each RREF matrix with leading ones in the positions  $j(i)$ . Using the product rule to fill these free positions one at a time with elements of  $K$ , we see that there are  $q^{|\mu|}$  RREF matrices with leading ones in the given positions. The theorem now follows from the sum rule, keeping in mind the bijection just constructed between  $j$ -sequences and partitions.  $\square$

**12.38. Example.** To illustrate the preceding proof, take  $n = 10$ ,  $k = 4$ , and consider RREF matrices of the form

$$\begin{bmatrix} 0 & 1 & * & * & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 1 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \end{bmatrix}.$$

Here  $*$ 's mark the free positions in the matrix, and  $(j(1), j(2), j(3), j(4)) = (2, 5, 6, 9)$ . The associated partition is  $\mu = (5, 3, 3, 1)$ , which does fit in a  $4 \times 6$  box. We can see the (reflected) diagram of this partition in the matrix by erasing the columns without stars and right-justifying the remaining columns. Evidently there are  $q^{12} = q^{|\mu|}$  ways of completing this template to get an RREF matrix with the leading ones in the indicated positions.

Going the other way, consider another partition  $\mu = (6, 2, 2, 0)$  that fits in a  $4 \times 6$  box. Using the formula  $j(i) = n - k + i - \mu_i$ , we recover  $(j(1), j(2), j(3), j(4)) = (1, 6, 7, 10)$ , which tells us the locations of the leading ones. So this particular partition corresponds to RREF matrices that match the following template:

$$\begin{bmatrix} 1 & * & * & * & * & 0 & 0 & * & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & * & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

## 12.8 Tangent and Secant Numbers

In calculus, one learns the following power series expansions for the trigonometric functions sine, cosine, and arctangent:

$$\sin x = x - x^3/3! + x^5/5! - x^7/7! + \dots = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!};$$

$$\cos x = 1 - x^2/2! + x^4/4! - x^6/6! - \dots = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!};$$

$$\arctan x = x - x^3/3 + x^5/5 - x^7/7 + \dots = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}.$$

These expansions are all special cases of Taylor's formula  $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$ . Using Taylor's formula, one can also find power series expansions for the tangent and secant functions:

$$\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \frac{17}{315}x^7 + \frac{62}{2835}x^9 + \frac{1382}{155925}x^{11} + \dots;$$

$$\sec x = 1 + \frac{1}{2}x^2 + \frac{5}{24}x^4 + \frac{61}{720}x^6 + \frac{277}{8064}x^8 + \frac{50521}{362880}x^{10} + \dots$$

The coefficients of these series seem quite irregular and unpredictable compared to the preceding three series. Remarkably, as we shall see in this section, these coefficients encode the solution to a counting problem involving permutations.

As in Chapter 7, we consider formal versions of the tangent and secant power series to avoid any questions of convergence. We define  $\tan x = \sin x / \cos x$  and  $\sec x = 1 / \cos x$ , where  $\sin x$  and  $\cos x$  are the formal series defined in 7.52. Now, for each  $n \geq 0$ , set  $a_n = (\tan x)^{(n)}(0)$  and  $b_n = (\sec x)^{(n)}(0)$ . The formal Maclaurin formula 7.55 asserts that

$$\tan x = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n; \quad \sec x = \sum_{n=0}^{\infty} \frac{b_n}{n!} x^n. \quad (12.1)$$

Since the ordinary Maclaurin series for the tangent and secant functions converge in a neighborhood of zero, the coefficients in the formal power series above match the coefficients in the ordinary power series representing the tangent and secant functions. The first several values of  $a_n$  and  $b_n$  are

$$\begin{aligned} (a_n : n \geq 0) &= (0, 1, 0, 2, 0, 16, 0, 272, 0, 7936, 0, 353792, \dots); \\ (b_n : n \geq 0) &= (1, 0, 1, 0, 5, 0, 61, 0, 1385, 0, 50521, \dots). \end{aligned} \quad (12.2)$$

One can check that  $a_n = 0$  for all even  $n$  and  $b_n = 0$  for all odd  $n$  (cf. 7.161).

Next, for each integer  $n \geq 0$ , let  $c_n$  be the number of permutations  $w = w_1 w_2 \cdots w_n$  of  $\{1, 2, \dots, n\}$  such that

$$w_1 < w_2 > w_3 < w_4 > \cdots < w_{n-1} > w_n; \quad (12.3)$$

note that  $c_n = 0$  for all even  $n$ . For each integer  $n \geq 0$ , let  $d_n$  be the number of permutations  $w$  of  $\{1, 2, \dots, n\}$  (or any  $n$ -letter ordered alphabet) such that

$$w_1 < w_2 > w_3 < w_4 > \cdots > w_{n-1} < w_n; \quad (12.4)$$

note that  $d_n = 0$  for all odd  $n$ . By reversing the ordering of the letters, one sees that  $d_n$  also counts the permutations  $w$  of  $n$  letters such that

$$w_1 > w_2 < w_3 > w_4 < \cdots < w_{n-1} > w_n. \quad (12.5)$$

Permutations of the form (12.3) or (12.4) are called *up-down permutations*. We aim to prove that  $a_n = c_n$  and  $b_n = d_n$  for all integers  $n \geq 0$ . The proof consists of five steps.

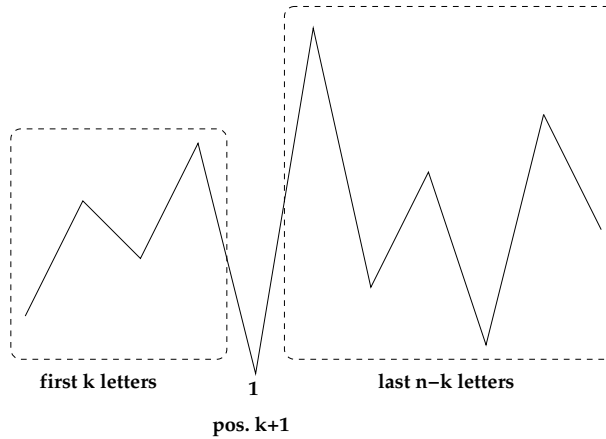
*Step 1.* Using the formal derivative rules, one may show that  $(\tan x)' = \sec^2 x$  (see 7.159). Differentiating the first series in (12.1), squaring the second series using 7.6, and equating the coefficients of  $x^n$ , we obtain

$$\frac{a_{n+1}}{n!} = \sum_{k=0}^n \frac{b_k}{k!} \frac{b_{n-k}}{(n-k)!}$$

or equivalently,

$$a_{n+1} = \sum_{k=0}^n \binom{n}{k} b_k b_{n-k}. \quad (12.6)$$

*Step 2.* We also have  $(\sec x)' = \tan x \sec x$  (see 7.159). Differentiating the second series



**FIGURE 12.12**

Counting up-down permutations of odd length.

in (12.1), multiplying the two series together using 7.6, and equating the coefficients of  $x^n$ , we obtain

$$\frac{b_{n+1}}{n!} = \sum_{k=0}^n \frac{a_k}{k!} \frac{b_{n-k}}{(n-k)!}$$

or equivalently,

$$b_{n+1} = \sum_{k=0}^n \binom{n}{k} a_k b_{n-k}. \quad (12.7)$$

*Step 3.* We give a counting argument to prove the relation

$$c_{n+1} = \sum_{k=0}^n \binom{n}{k} d_k d_{n-k}. \quad (12.8)$$

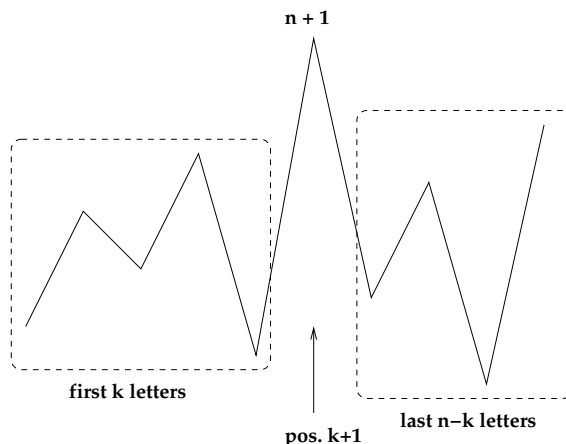
If  $n$  is odd, then both sides of this relation are zero, since at least one of  $k$  or  $n - k$  is odd for each  $k$ . Now suppose  $n$  is even. How can we build a typical permutation

$$w = w_1 < w_2 > w_3 < \cdots > w_{n+1}$$

counted by  $c_{n+1}$ ? Let us first choose the position of 1 in  $w$ ; say  $w_{k+1} = 1$  for some  $k$  between 0 and  $n$ . The required inequalities at position  $k + 1$  will be satisfied if and only if  $k$  is even. Observe that in the case where  $k$  is odd,  $d_k d_{n-k} = 0$  so this term contributes nothing to the right side of (12.8). Given that  $k$  is even, choose a  $k$ -element subset  $A$  of the  $n$  remaining letters in  $\binom{n}{k}$  ways. Use these letters to fill in the first  $k$  positions of  $w$ , subject to the required inequalities (12.4), in any of  $d_k$  ways. Use the remaining letters to fill in the last  $(n + 1) - (k + 1) = n - k$  positions of  $w$  (subject to the inequalities (12.5), reindexed to begin at index  $k + 2$ ), in any of  $d_{n-k}$  ways. The desired relation now follows from the sum and product rules. See Figure 12.12, in which  $w$  is visualized as a sequence of line segments connecting the points  $(i, w_i)$  for  $1 \leq i \leq n + 1$ .

*Step 4.* We give a counting argument to prove the relation

$$d_{n+1} = \sum_{k=0}^n \binom{n}{k} c_k d_{n-k}. \quad (12.9)$$

**FIGURE 12.13**

Counting up-down permutations of even length.

Both sides are zero if  $n$  is even. If  $n$  is odd, we must build a permutation

$$w = w_1 < w_2 > w_3 < \cdots < w_{n+1}.$$

First choose an index  $k$  with  $0 \leq k \leq n$ , and define  $w_{k+1} = n+1$ . This time, to get a nonzero contribution from this value of  $k$ , we need  $k$  to be odd. Now pick a  $k$ -element subset  $A$  of the  $n$  remaining letters. Use the letters in  $A$  to fill in  $w_1, w_2, \dots, w_k$  ( $c_k$  ways), and use the remaining letters to fill in  $w_{k+2}, \dots, w_{n+1}$  ( $d_{n-k}$  ways). See Figure 12.13.

*Step 5:* A routine induction argument now shows that  $a_n = c_n$  and  $b_n = d_n$  for all  $n \geq 0$ , since the pair of sequences  $(a_n), (b_n)$  satisfy the same system of recursions and initial conditions as the pair of sequences  $(c_n), (d_n)$ . This completes the proof.

## 12.9 Tournaments and the Vandermonde Determinant

This section uses the combinatorics of tournaments to prove a famous determinant formula.

**12.39. Definition: Tournaments.** An  $n$ -player tournament is a digraph  $t$  with vertex set  $[n] = \{1, 2, \dots, n\}$  such that, for  $1 \leq i < j \leq n$ , exactly one of the directed edges  $(i, j)$  or  $(j, i)$  is an edge of  $t$ . Let  $T_n$  be the set of all such tournaments.

Intuitively, the  $n$  vertices represent  $n$  players who compete in a series of one-on-one matches. Each player plays every other player exactly once, and there are no ties. If player  $i$  beats player  $j$ , the edge  $(i, j)$  is part of the tournament; otherwise, the edge  $(j, i)$  is included.

**12.40. Definition: Weights, Inversions, and Sign for Tournaments.** Suppose  $t \in T_n$  is a tournament. The *weight* of  $t$  is  $\text{wt}(t) = \prod_{i=1}^n x_i^{\text{outdeg}_t(i)}$ . The *inversion number* of  $t$  is  $\text{inv}(t) = \sum_{1 \leq i < j \leq n} \chi((j, i) \in t)$ . The *sign* of  $t$  is  $\text{sgn}(t) = (-1)^{\text{inv}(t)}$ .

Informally,  $\text{wt}(t) = x_1^{e_1} \cdots x_n^{e_n}$  iff player  $i$  beats  $e_i$  other players for all  $i$ . If we think of the numbers  $1, 2, \dots, n$  as giving the initial rankings of the players,  $\text{inv}(t)$  counts the number of times a lower-ranked player beats a higher-ranked one (with 1 being the highest rank).

**12.41. Example.** Consider the tournament  $t \in T_5$  with edge set

$$\{(1, 3), (1, 4), (1, 5), (2, 1), (2, 4), (3, 2), (3, 4), (3, 5), (5, 2), (5, 4)\}.$$

We have  $\text{wt}(t) = x_1^3 x_2^2 x_3^3 x_4^2$ ,  $\text{inv}(t) = 4$ , and  $\text{sgn}(t) = +1$ .

**12.42. Theorem: Tournament Generating Function.** For all  $n \geq 1$ ,

$$\sum_{t \in T_n} \text{sgn}(t) \text{wt}(t) = \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

*Proof.* We can build a tournament  $t \in T_n$  by making a sequence of binary choices, indexed by the pairs  $i < j$  with  $i, j \in [n]$ : for each  $i < j$ , pick either  $(i, j)$  or  $(j, i)$  and add this edge to  $t$ . Let us examine the effect of this choice on  $\text{wt}(t)$ ,  $\text{inv}(t)$ , and  $\text{sgn}(t)$ . If we add  $(i, j)$  to  $t$  (so  $i$  beats  $j$ ), the exponent of  $x_i$  goes up by 1, inversions go up by zero, and the sign is unchanged. If we add  $(j, i)$  to  $t$  instead, the exponent of  $x_j$  goes up by 1, inversions go up by one, and the sign is multiplied by  $-1$ . The generating function  $(+x_i - x_j)$  records the effect of this choice. The proof is completed by invoking the product rule for generating functions.  $\square$

Given a tournament  $t$ , there may exist three players  $u, v, w$  where  $u$  beats  $v$ ,  $v$  beats  $w$ , and  $w$  beats  $u$ . This situation occurs whenever the digraph  $t$  contains a directed 3-cycle. Let us give a name to tournaments where this circularity condition does *not* occur.

**12.43. Definition: Transitive Tournaments.** A tournament  $t \in T_n$  is *transitive* iff for all  $u, v, w \in [n]$ ,  $(u, v) \in t$  and  $(v, w) \in t$  imply  $(u, w) \in t$ .

Note that  $(u, v) \in t$  and  $(v, w) \in t$  force  $u \neq w$ , and then  $(u, w) \notin t$  is equivalent to  $(w, u) \in t$ . It follows that a tournament is not transitive iff there exist  $u, v, w \in [n]$  with  $(u, v) \in t$  and  $(v, w) \in t$  and  $(w, u) \in t$ .

**12.44. Theorem: Generating Function for Transitive Tournaments.** Let  $T'_n$  be the set of transitive tournaments in  $T_n$ . Then

$$\sum_{t \in T'_n} \text{sgn}(t) \text{wt}(t) = \sum_{w \in S_n} \text{sgn}(w) \prod_{k=1}^n x_{w(k)}^{n-k}.$$

*Proof.* We define a bijection  $f : T'_n \rightarrow S_n$  that will be used to transfer signs and weights from  $T'_n$  to  $S_n$ . Given  $t \in T'_n$ , define an associated relation  $\preceq$  on  $[n]$  by setting  $u \preceq v$  iff  $u = v$  or  $(u, v) \in t$ . This relation is evidently reflexive, antisymmetric (since  $t$  is a tournament), and transitive (since  $t$  is transitive). Furthermore,  $u \preceq v$  or  $v \preceq u$  for all  $u, v \in [n]$  since  $t$  is a tournament. So  $\preceq$  is a total ordering of  $[n]$ . This ordering determines a unique permutation of the players, namely

$$f(t) = w = w_1 \prec w_2 \prec \cdots \prec w_n.$$

For all  $k$ , player  $w_k$  beats all players  $w_m$  for  $m > k$  and loses to all players  $w_m$  for  $m < k$ . This remark shows that  $t$  is uniquely determined by  $w$ , so the map  $f$  is a bijection.

Let us compare  $\text{inv}(t)$  to  $\text{inv}(w)$ , where  $w = f(t)$ . Consider two players  $i = w_k$  and  $j = w_m$  with  $i < j$  (so  $w_m > w_k$ ). This pair contributes to  $\text{inv}(t)$  iff  $(j, i) \in t$  iff  $j$  beats  $i$  in  $t$  iff  $j$  appears before  $i$  in  $w$  iff  $m < k$  iff the letters in positions  $m, k$  of  $w$  contribute to  $\text{inv}(w)$ . So  $\text{inv}(t) = \text{inv}(w)$  and  $\text{sgn}(t) = \text{sgn}(w)$ . Next, let us express  $\text{wt}(t)$  in terms of  $w$ . Since player  $w_k$  beats all players in the range  $k < m \leq n$ , we see that

$$\text{wt}(t) = \prod_{k=1}^n x_{w_k}^{n-k}.$$



Define  $\text{wt}(w)$  by the right side of this formula. The theorem now follows because  $f$  is a weight-preserving, sign-preserving bijection.  $\square$

We can use the bijection  $f$  to characterize transitive tournaments.

**12.45. Theorem: Criterion for Transitive Tournaments.** A tournament  $t \in T_n$  is transitive iff no two vertices in  $[n]$  have the same outdegree.

*Proof.* If  $t$  is transitive, consider  $w = f(t) = w_1 \prec w_2 \prec \cdots \prec w_n$ . As shown above,  $\text{wt}(t) = \prod_{k=1}^n x_{w_k}^{n-k}$ . The exponents  $n-k$  are all distinct, so every vertex has a different outdegree. Conversely, suppose  $t \in T_n$  is such that every vertex has a different outdegree. There are  $n$  vertices and  $n$  possible outdegrees (namely  $0, 1, \dots, n-1$ ), so each possible outdegree occurs at exactly one vertex. Let  $w_1$  be the unique vertex with outdegree  $n-1$ . Then  $w_1$  beats all other players. Next, let  $w_2$  be the unique vertex with outdegree  $n-2$ . Then  $w_2$  must beat all players except  $w_1$ . Continuing similarly, we obtain a permutation  $w = w_1, w_2, \dots, w_n$  of  $[n]$  such that  $w_j$  beats  $w_k$  iff  $j < k$ . To confirm that  $t$  is transitive, consider three players  $w_i, w_j, w_k$  with  $(w_i, w_j) \in t$  and  $(w_j, w_k) \in t$ . Then  $i < j$  and  $j < k$ , so  $i < k$ , so  $(w_i, w_k) \in t$ .  $\square$

**12.46. Theorem: Vandermonde Determinant Formula.** Let  $x_1, \dots, x_n$  be fixed elements in a commutative ring  $R$ . Define an  $n \times n$  matrix  $V$  by setting  $V(i, j) = x_j^{n-i}$  for  $1 \leq i, j \leq n$ . Then

$$\det(V) = \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

*Proof.* According to the definition of determinants in 9.37,

$$\det(V) = \sum_{w \in S_n} \text{sgn}(w) \prod_{k=1}^n V(k, w(k)) = \sum_{w \in S_n} \text{sgn}(w) \prod_{k=1}^n x_{w(k)}^{n-k}. \quad (12.10)$$

This is the generating function for transitive tournaments, whereas  $\prod_{i < j} (x_i - x_j)$  is the generating function for all tournaments with  $n$  players. So, it suffices to define a sign-reversing, weight-preserving involution  $I : T_n \rightarrow T_n$  with fixed point set  $T'_n$ . Define  $I(t) = t$  for  $t \in T'_n$ . Now consider a non-transitive tournament  $t \in T_n \sim T'_n$ . By 12.45, there exist two vertices  $i < j \in [n]$  with the same outdegree in  $t$ . If there are several pairs of vertices with the same outdegree, choose the pair such that  $i$  and then  $j$  is minimized. Define  $I(t)$  by switching the roles of  $i$  and  $j$  in  $t$ ; more precisely, replace every directed edge  $(u, v)$  in  $t$  by  $(s_{i,j}(u), s_{i,j}(v))$ , where  $s_{i,j}$  is the transposition  $(i, j) \in S_n$ . The resulting tournament is non-transitive (since  $i$  and  $j$  still have the same outdegree in  $I(t)$ ) and has the same weight as  $t$ . Furthermore,  $I(I(t)) = t$ .

Finally, we show that  $\text{sgn}(I(t)) = -\text{sgn}(t)$ . Consider the factorization of  $(i, j) \in S_n$  into  $2(j-i)-1$  basic transpositions:

$$(i, j) = (j-1, j)(j-2, j-1) \cdots (i+1, i+2)(i, i+1)(i+1, i+2) \cdots (j-2, j-1)(j-1, j).$$

We can pass from  $t$  to  $I(t)$  in stages, by applying these basic transpositions one at a time to the endpoints of the directed edges in  $t$ . We claim that each such step changes the sign of the tournament. For, consider what happens to the inversion count when we pass from a tournament  $z$  to  $z'$  by switching labels  $k$  and  $k+1$ . The inversion  $(k+1, k)$  is present in exactly one of the tournaments  $z$  and  $z'$ , and the other inversions are unaffected by the label switch. So  $\text{inv}(z')$  differs from  $\text{inv}(z)$  by  $\pm 1$ , and hence  $\text{sgn}(z') = -\text{sgn}(z)$ . Since we pass from  $t$  to  $I(t)$  by an odd number of moves of this type (namely  $2(j-i)-1$ ), we see that  $\text{sgn}(I(t)) = -\text{sgn}(t)$ , as desired.  $\square$

## 12.10 Hook-Length Formula

This section presents a probabilistic proof of the hook-length formula for the number of standard tableaux of a given shape. This formula was first stated in the Introduction. For the reader's convenience, we begin by recalling the relevant definitions.

**12.47. Definitions.** An *integer partition of  $n$*  is a weakly decreasing sequence  $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l)$  of positive integers with  $\lambda_1 + \cdots + \lambda_l = n$ . The *diagram* of  $\lambda$  is

$$\text{dg}(\lambda) = \{(i, j) \in \mathbb{N} \times \mathbb{N} : 1 \leq i \leq l, 1 \leq j \leq \lambda_i\}.$$

Each  $(i, j) \in \text{dg}(\lambda)$  is called a *box* or a *cell*. We take  $i$  as the row index and  $j$  as the column index, where the topmost row is row 1. Given any cell  $c = (i, j) \in \text{dg}(\lambda)$ , the *hook* of  $c$  in  $\lambda$  is

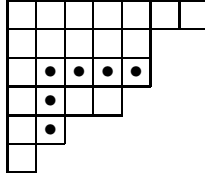
$$H(c) = \{(i, k) \in \text{dg}(\lambda) : k \geq j\} \cup \{(k, j) \in \text{dg}(\lambda) : k \geq i\}.$$

The *hook-length* of  $c$  in  $\lambda$  is  $h(c) = |H(c)|$ . A *corner box* of  $\lambda$  is a cell  $c \in \text{dg}(\lambda)$  with  $h(c) = 1$ . A *standard tableau of shape  $\lambda$*  is a bijection  $S : \text{dg}(\lambda) \rightarrow \{1, 2, \dots, n\}$  such that  $S(i, j) < S(i, j+1)$  for all  $i, j$  such that  $(i, j), (i, j+1) \in \text{dg}(\lambda)$ , and  $S(i, j) < S(i+1, j)$  for all  $i, j$  such that  $(i, j), (i+1, j) \in \text{dg}(\lambda)$ . Let  $\text{SYT}(\lambda)$  be the set of standard tableaux of shape  $\lambda$ , and let  $f^\lambda = |\text{SYT}(\lambda)|$ .

**12.48. Example.** If  $\lambda = (7, 5, 5, 4, 2, 1)$  and  $c = (3, 2)$ , then

$$H(c) = \{(3, 2), (3, 3), (3, 4), (3, 5), (4, 2), (5, 2)\}$$

and  $h(c) = 6$ . We can visualize  $\text{dg}(\lambda)$  and  $H(c)$  using the following picture.



Let  $\lambda'_j$  be the number of boxes in column  $j$  of  $\text{dg}(\lambda)$ . Then  $h(i, j) = (\lambda_i - j) + (\lambda'_j - i) + 1$ . We use this formula to establish the following lemma.

**12.49. Lemma.** Suppose  $\lambda$  is a partition of  $n$ ,  $(r, s)$  is a corner box of  $\lambda$ , and  $(i, j) \in \text{dg}(\lambda)$  satisfies  $i < r$  and  $j < s$ . Then  $h(i, j) = h(r, j) + h(i, s) - 1$ .

*Proof.* Since  $(r, s)$  is a corner box,  $\lambda_r = s$  and  $\lambda'_s = r$ . So

$$\begin{aligned} h(r, j) + h(i, s) - 1 &= [(\lambda_r - j) + (\lambda'_j - r) + 1] + [(\lambda_i - s) + (\lambda'_s - i) + 1] - 1 \\ &= s - j + \lambda'_j - r + \lambda_i - s + r - i + 1 \\ &= (\lambda_i - j) + (\lambda'_j - i) + 1 = h(i, j). \quad \square \end{aligned}$$

**12.50. Theorem: Hook-Length Formula.** For any partition  $\lambda$  of  $n$ ,

$$f^\lambda = \frac{n!}{\prod_{c \in \text{dg}(\lambda)} h(c)}.$$

The idea of the proof is to define a *random algorithm* that takes a partition  $\lambda$  of  $n$  as input and produces a standard tableau  $S \in \text{SYT}(\lambda)$  as output. We will prove in 12.55 that this algorithm outputs any given standard tableau  $S$  with probability

$$p = \frac{\prod_{c \in \text{dg}(\lambda)} h(c)}{n!}.$$

This probability depends only on  $\lambda$ , not on  $S$ , so we obtain a uniform probability distribution on the sample space  $\text{SYT}(\lambda)$ . So, on one hand, each standard tableau is produced with probability  $p$ ; and on the other hand, each standard tableau is produced with probability  $1/|\text{SYT}(\lambda)| = 1/f^\lambda$ . Thus  $f^\lambda = 1/p$ , and we obtain the hook-length formula.

Here is an informal description of the algorithm for generating a random standard tableau of shape  $\lambda$ . Start at a random cell in the shape  $\lambda$ . As long as we are not at a corner box, we jump from our current box  $c$  to some other cell in  $H(c)$ ; each cell in the hook is chosen with equal probability. This jumping process eventually takes us to a corner cell. We place the entry  $n$  in this box, and then pretend this cell is no longer there. We are left with a partition  $\mu$  of size  $n - 1$ . Proceed recursively to select a random standard tableau of shape  $\mu$ . Adding back the corner cell containing  $n$  gives the desired tableau of shape  $\lambda$ .

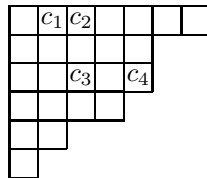
Now we give a formal description of the algorithm. Every random choice below is to be independent of all other choices.

**12.51. Tableau Generation Algorithm.** The input to the algorithm is a partition  $\lambda$  of  $n$ . The output is a tableau  $S \in \text{SYT}(\lambda)$ , constructed according to the following random procedure. As a base case, if  $n = 0$ , return the empty tableau of shape 0.

1. Choose a random cell  $c \in \text{dg}(\lambda)$ . Each cell in  $\text{dg}(\lambda)$  is chosen with probability  $1/n$ .
2. While  $h(c) > 1$ , do the following.
  - 2a. Choose a random cell  $c' \in H(c) \sim \{c\}$ . Each cell in  $H(c) \sim \{c\}$  is chosen with probability  $1/(h(c) - 1)$ .
  - 2b. Replace  $c$  by  $c'$  and go back to step 2.
3. Now  $c$  is a corner box of  $\text{dg}(\lambda)$ , so  $\text{dg}(\lambda) \sim \{c\}$  is the diagram of some partition  $\mu$  of  $n - 1$ . Recursively use the same algorithm to generate a random standard tableau  $S' \in \text{SYT}(\mu)$ . Extend this to a standard tableau  $S \in \text{SYT}(\lambda)$  by setting  $S(c) = n$ , and output  $S$  as the answer.

Let  $(c_1, c_2, c_3, \dots, c_k)$  be the sequence of cells chosen in steps 1 and 2. Call this sequence the *hook walk for  $n$* . Note that the hook walk must be finite, since  $h(c_1) > h(c_2) > h(c_3) > \dots$ . Writing  $c_s = (i_s, j_s)$  for each  $s$ , define  $I = \{i_1, \dots, i_{k-1}\} \sim \{i_k\}$  and  $J = \{j_1, \dots, j_{k-1}\} \sim \{j_k\}$ . We call  $I$  and  $J$  the *row set* and *column set* for this hook walk.

**12.52. Example.** Given  $n = 24$  and  $\lambda = (7, 5, 5, 4, 2, 1)$ , the first iteration of the algorithm might proceed as follows.



Here we will place  $n = 24$  in corner box  $c_4$  and proceed recursively to fill in the rest of the tableau. The probability that the algorithm will choose this particular hook walk for  $n$  is

$$\frac{1}{n} \cdot \frac{1}{h(c_1) - 1} \cdot \frac{1}{h(c_2) - 1} \cdot \frac{1}{h(c_3) - 1} = \frac{1}{24} \cdot \frac{1}{9} \cdot \frac{1}{7} \cdot \frac{1}{3}.$$

The row set and column set for this hook walk are  $I = \{1\}$  and  $J = \{2, 3\}$ .

The next lemma is the key technical fact needed to analyze the behaviour of the tableau generation algorithm.

**12.53. Lemma.** Given a partition  $\lambda$  of  $n$ , a corner box  $c = (r, s)$ , and sets  $I \subseteq \{1, 2, \dots, r-1\}$  and  $J \subseteq \{1, 2, \dots, s-1\}$ , the probability that the hook walk for  $n$  ends at  $c$  with row set  $I$  and column set  $J$  is

$$p(\lambda, c, I, J) = \frac{1}{n} \prod_{i \in I} \frac{1}{h(i, s) - 1} \prod_{j \in J} \frac{1}{h(r, j) - 1}.$$

*Proof.* Write  $I = \{i_1 < i_2 < \dots < i_\ell\}$  and  $J = \{j_1 < j_2 < \dots < j_m\}$ , where  $\ell, m \geq 0$ . First we consider some degenerate cases. Say  $I = J = \emptyset$ . Then the hook walk for  $n$  consists of the single cell  $c$ . This happens with probability  $1/n$ , in agreement with the formula in the lemma (interpreting the empty products as 1). Next, suppose  $I$  is empty but  $J$  is not. The hook walk for  $n$  in this case must be  $c_1 = (r, j_1)$ ,  $c_2 = (r, j_2)$ ,  $\dots$ ,  $c_m = (r, j_m)$ ,  $c_{m+1} = (r, s)$ . The probability of this hook walk is

$$\frac{1}{n} \cdot \frac{1}{h(c_1) - 1} \cdot \frac{1}{h(c_2) - 1} \cdots \frac{1}{h(c_m) - 1} = \frac{1}{n} \prod_{j \in J} \frac{1}{h(r, j) - 1}.$$

Similarly, the result holds when  $J$  is empty and  $I$  is nonempty.

Now consider the case where both  $I$  and  $J$  are nonempty. We will argue by induction on  $|I| + |J|$ . A hook walk with row set  $I$  and column set  $J$  ending at  $c$  must begin with the cell  $c_1 = (i_1, j_1)$ ; this cell is chosen in step 1 of the algorithm with probability  $1/n$ . Now, there are two possibilities for cell  $c_2$ : either  $c_2 = (i_1, j_2)$  or  $c_2 = (i_2, j_1)$ . Each possibility for  $c_2$  is chosen with probability  $1/(h(c) - 1) = 1/(h(i_1, j_1) - 1)$ . When  $c_2 = (i_1, j_2)$ , the sequence  $(c_2, \dots, c_k)$  is a hook walk ending at  $c$  with row set  $I$  and column set  $J' = J \sim \{j_1\}$ . By induction, such a hook walk occurs with probability

$$\frac{1}{n} \prod_{i \in I} \frac{1}{h(i, s) - 1} \prod_{j \in J'} \frac{1}{h(r, j) - 1}.$$

However, since the walk really started at  $c_1$  and proceeded to  $c_2$ , we replace the first factor  $1/n$  by  $\frac{1}{n} \cdot \frac{1}{h(c_1) - 1}$ . Similarly, when  $c_2 = (i_2, j_1)$ , the sequence  $(c_2, \dots, c_k)$  is a hook walk ending at  $c$  with row set  $I' = I \sim \{i_1\}$  and column set  $J$ . So the probability that the hook walk starts at  $c_1$  and proceeds through  $c_2 = (i_2, j_1)$  is

$$\frac{1}{n} \cdot \frac{1}{h(c_1) - 1} \prod_{i \in I'} \frac{1}{h(i, s) - 1} \prod_{j \in J} \frac{1}{h(r, j) - 1}.$$

Adding these two terms, we see that

$$p(\lambda, c, I, J) = \frac{1}{n} \cdot \frac{1}{h(c_1) - 1} \prod_{i \in I'} \frac{1}{h(i, s) - 1} \prod_{j \in J'} \frac{1}{h(r, j) - 1} \cdot \left( \frac{1}{h(i_1, s) - 1} + \frac{1}{h(r, j_1) - 1} \right).$$

The factor in parentheses is

$$\frac{h(r, j_1) + h(i_1, s) - 2}{(h(i_1, s) - 1)(h(r, j_1) - 1)}.$$

Using 12.49, the numerator simplifies to  $h(i_1, j_1) - 1 = h(c_1) - 1$ . This factor cancels and leaves us with

$$p(\lambda, c, I, J) = \frac{1}{n} \prod_{i \in I} \frac{1}{h(i, s) - 1} \prod_{j \in J} \frac{1}{h(r, j) - 1}.$$

This completes the induction proof.  $\square$

**12.54. Theorem: Probability that a Hook Walk ends at  $c$ .** Given a partition  $\lambda$  of  $n$  and a corner box  $c = (r, s)$  of  $\text{dg}(\lambda)$ , the probability that the hook walk for  $n$  ends at  $c$  is

$$p(\lambda, c) = \frac{1}{n} \prod_{i=1}^{r-1} \frac{h(i, s)}{h(i, s) - 1} \prod_{j=1}^{s-1} \frac{h(r, j)}{h(r, j) - 1}.$$

*Proof.* Write  $[r-1] = \{1, 2, \dots, r-1\}$  and  $[s-1] = \{1, 2, \dots, s-1\}$ . By the sum rule for probabilities,

$$\begin{aligned} p(\lambda, c) &= \sum_{I \subseteq [r-1]} \sum_{J \subseteq [s-1]} p(\lambda, c, I, J) \\ &= \frac{1}{n} \sum_{I \subseteq [r-1]} \sum_{J \subseteq [s-1]} \prod_{i \in I} \frac{1}{h(i, s) - 1} \prod_{j \in J} \frac{1}{h(r, j) - 1} \\ &= \frac{1}{n} \left( \sum_{I \subseteq [r-1]} \prod_{i \in I} \frac{1}{h(i, s) - 1} \right) \cdot \left( \sum_{J \subseteq [s-1]} \prod_{j \in J} \frac{1}{h(r, j) - 1} \right). \end{aligned}$$

By 2.7, we have

$$\sum_{I \subseteq [r-1]} \prod_{i \in I} \frac{1}{h(i, s) - 1} = \prod_{i=1}^{r-1} \left( 1 + \frac{1}{h(i, s) - 1} \right) = \prod_{i=1}^{r-1} \frac{h(i, s)}{h(i, s) - 1}.$$

The sum over  $J$  can be simplified in a similar way, giving the formula in the theorem.  $\square$

The next theorem is the final step in the proof of the hook-length formula.

**12.55. Theorem: Probability of Generating a Given Tableau.** If  $\lambda$  is a partition of  $n$  and  $S \in \text{SYT}(\lambda)$ , the tableau generation algorithm outputs  $S$  with probability

$$\frac{\prod_{c \in \text{dg}(\lambda)} h(c)}{n!}.$$

*Proof.* We prove the theorem by induction on  $n$ . Note first that the result does hold for  $n = 0$  and  $n = 1$ . For the induction step, assume the result is known for partitions and tableaux with fewer than  $n$  boxes. Let  $c^* = (r, s)$  be the cell such that  $S(c^*) = n$ , let  $\mu$  be the partition obtained by removing  $c^*$  from  $\text{dg}(\lambda)$ , and let  $S' \in \text{SYT}(\mu)$  be the tableau obtained by erasing  $n$  from  $S$ . First, the probability that the hook walk for  $n$  (in steps 1 and 2 of the algorithm) ends at  $c^*$  is  $p(\lambda, c^*)$ . Given that this has occurred, induction tells us that the probability of generating  $S'$  in step 3 is

$$\frac{\prod_{c \in \text{dg}(\mu)} h_\mu(c)}{(n-1)!},$$

where  $h_\mu(c)$  refers to the hook length of  $c$  relative to  $\text{dg}(\mu)$ . Multiplying these probabilities, the probability of generating  $S$  is therefore

$$\frac{1}{n!} \prod_{c \in \text{dg}(\mu)} h_\mu(c) \prod_{i=1}^{r-1} \frac{h_\lambda(i, s)}{h_\lambda(i, s) - 1} \prod_{j=1}^{s-1} \frac{h_\lambda(r, j)}{h_\lambda(r, j) - 1}.$$

Now, consider what happens to the hook lengths of cells when we pass from  $\mu$  to  $\lambda$  by restoring the box  $c^* = (r, s)$ . For every cell  $c = (i, j) \in \text{dg}(\mu)$  with  $i \neq r$  and  $j \neq s$ , we

have  $h_\mu(c) = h_\lambda(c)$ . If  $c = (i, s) \in \text{dg}(\mu)$  with  $i < r$ , then  $h_\mu(c) = h_\lambda(c) - 1 = h_\lambda(i, s) - 1$ . Thus, the fractions in the second product convert  $h_\mu(c)$  to  $h_\lambda(c)$  for each such  $c$ . Similarly, if  $c = (r, j) \in \text{dg}(\mu)$  with  $j < s$ , then  $h_\mu(c) = h_\lambda(c) - 1 = h_\lambda(r, j) - 1$ . So the fractions in the third product convert  $h_\mu(c)$  to  $h_\lambda(c)$  for each such  $c$ . So we are left with

$$\frac{1}{n!} \prod_{c \in \text{dg}(\mu)} h_\lambda(c) = \frac{\prod_{c \in \text{dg}(\lambda)} h_\lambda(c)}{n!},$$

where the last equality follows since  $h_\lambda(c^*) = 1$ . This completes the induction.  $\square$

## 12.11 Knuth Equivalence

Let  $X$  be a totally ordered set, and let  $X^* = \bigcup_{n \geq 0} X^n$  be the set of all words over the alphabet  $X$ . Given a word  $w \in X^*$ , we can use the RSK algorithm to construct the insertion tableau  $P(w)$ , which is a semistandard tableau using the same multiset of letters as  $w$  (§10.23). This section studies some of the relationships between  $w$  and  $P(w)$ . In particular, we show that the shape of  $P(w)$  contains information about increasing and decreasing subsequences of  $w$ . First we show how to encode semistandard tableaux using words.

**12.56. Definition: Reading Word of a Tableau.** Let  $T \in \text{SSYT}_X(\lambda)$ , with  $\lambda = (\lambda_1, \dots, \lambda_k)$ . The *reading word* of  $T$  is

$$\text{rw}(T) = T(k, 1), T(k, 2), \dots, T(k, \lambda_k), T(k-1, 1), T(k-1, 2), \dots, T(k-1, \lambda_{k-1}), \dots, T(1, 1), T(1, 2), \dots, T(1, \lambda_1).$$

Thus,  $\text{rw}(T)$  is the concatenation of the weakly increasing words appearing in each row of  $T$ , reading the rows from bottom to top. Note that  $T(j, \lambda_j) \geq T(j, 1) > T(j-1, 1)$  for all  $j > 1$ . This implies that we can recover the shape of  $T$  from  $\text{rw}(T)$  by starting a new row whenever we see a strict descent in  $\text{rw}(T)$ .

**12.57. Example.** Given the tableau

$$T = \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 2 & 3 & 4 & 4 & 6 \\ \hline 2 & 4 & 5 & 6 & 6 & & \\ \hline 3 & 5 & 7 & 8 & & & \\ \hline 4 & 6 & & & & & \\ \hline \end{array},$$

the reading word of  $T$  is

$$\text{rw}(T) = 463578245661123446.$$

Given that the word  $w = 7866453446223511224$  is the reading word of some tableau  $S$ , we deduce that  $S$  must be

$$S = \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 2 & 2 & 4 \\ \hline 2 & 2 & 3 & 5 & \\ \hline 3 & 4 & 4 & 6 & \\ \hline 4 & 5 & & & \\ \hline 6 & 6 & & & \\ \hline 7 & 8 & & & \\ \hline \end{array}$$

by looking at the descents in  $w$ .

Next we introduce two equivalence relations on  $X^*$  that are related to the map  $w \mapsto P(w)$ .

**12.58. Definition:  $P$ -Equivalence.** Two words  $v, w \in X^*$  are called  $P$ -equivalent, denoted  $v \equiv_P w$ , iff  $P(v) = P(w)$ .

**12.59. Definition: Knuth Equivalence.** The set of *elementary Knuth relations of the first kind* on  $X$  is

$$K_1 = \{(uyxz\mathbf{v}, \mathbf{u}yzx\mathbf{v}) : \mathbf{u}, \mathbf{v} \in X^*, x, y, z \in X \text{ and } x < y \leq z\}.$$

The set of *elementary Knuth relations of the second kind* on  $X$  is

$$K_2 = \{(\mathbf{u}xzy\mathbf{v}, \mathbf{u}zxy\mathbf{v}) : \mathbf{u}, \mathbf{v} \in X^*, x, y, z \in X \text{ and } x \leq y < z\}.$$

Two words  $v, w \in X^*$  are *Knuth equivalent*, denoted  $v \equiv_K w$ , iff there is a finite sequence of words  $v = v^0, v^1, v^2, \dots, v^k = w$  such that, for  $1 \leq i \leq k$ , either  $(v^{i-1}, v^i) \in K_1 \cup K_2$  or  $(v^i, v^{i-1}) \in K_1 \cup K_2$ .

**12.60. Remark.** Informally, Knuth equivalence allows us to modify words by repeatedly changing subsequences of three consecutive letters according to certain rules. Specifically, if the middle *value* among the three letters does not occupy the middle *position*, then the other two values can switch positions. To determine which value is the “middle value” in the case of repeated letters, use the rule that the letter to the right is larger. These comments should aid the reader in remembering the inequalities in the definitions of  $K_1$  and  $K_2$ .

It is routine to check that  $\equiv_P$  and  $\equiv_K$  are equivalence relations on  $X^*$ . Our current goal is to prove that these equivalence relations are actually the *same*. First we show that we can simulate each step in the tableau insertion algorithm 10.52 using the elementary Knuth relations.

**12.61. Theorem: Reading Words and Knuth Equivalence.** For all  $v \in X^*$ ,  $v \equiv_K \text{rw}(P(v))$ .

*Proof.* First note that, for any words  $u, z, w, w' \in X^*$ , if  $w \equiv_K w'$  then  $uwz \equiv_K uw'z$ . Now, write  $v = v_1v_2 \cdots v_k$  and argue by induction on  $k$ . The theorem holds if  $k \leq 1$ , since  $\text{rw}(P(v)) = v$  in this case. For the induction step, assume  $k > 1$  and write  $T' = P(v_1v_2 \cdots v_{k-1})$ ,  $T = P(v)$ . By the induction hypothesis,  $v_1 \cdots v_{k-1} \equiv_K \text{rw}(T')$ , so  $v = (v_1 \cdots v_{k-1})v_k \equiv_K \text{rw}(T')v_k$ . It will therefore suffice to prove that  $\text{rw}(T')v_k$  is Knuth equivalent to  $\text{rw}(T) = \text{rw}(T' \leftarrow v_k)$ . This will be proved by induction on  $\ell$ , the number of rows in the tableau  $T'$ .

For the base case, let  $\ell = 1$ . Then  $\text{rw}(T')$  is a weakly increasing sequence  $u_1u_2 \cdots u_{k-1}$ . If  $u_{k-1} \leq v_k$ , then  $T$  is obtained from  $T'$  by appending  $v_k$  at the end of the first row. In this situation,  $\text{rw}(T')v_k = u_1 \cdots u_{k-1}v_k = \text{rw}(T)$ , so the desired result holds. On the other hand, if  $v_k < u_{k-1}$ , let  $j$  be the least index with  $v_k < u_j$ . When inserting  $v_k$  into  $T'$ ,  $v_k$  will bump  $u_j$  into the second row, so that

$$\text{rw}(T) = u_ju_1u_2 \cdots u_{j-1}v_ku_{j+1} \cdots u_{k-1}.$$

Let us show that this word can be obtained from  $u_1 \cdots u_{k-1}v_k$  by a sequence of elementary Knuth equivalences. If  $j \leq k-2$ , then  $v_k < u_{k-2} \leq u_{k-1}$  implies

$$(u_1 \cdots u_{k-3}u_{k-2}v_ku_{k-1}, u_1 \cdots u_{k-3}u_{k-2}u_{k-1}v_k) \in K_1.$$

So  $\text{rw}(T')v_k$  is Knuth-equivalent to the word obtained by interchanging  $v_k$  with the letter  $u_{k-1}$  to its immediate left. Similarly, if  $j \leq k-3$ , the inequality  $v_k < u_{k-3} \leq u_{k-2}$  lets us

interchange  $v_k$  with  $u_{k-2}$ . We can continue in this way, using elementary Knuth relations of the first kind, to see that

$$\text{rw}(T')v_k \equiv_K u_1 \cdots u_{j-1} u_j v_k u_{j+1} \cdots u_{k-1}.$$

Now, we have  $u_{j-1} \leq v_k < u_j$ , so an elementary Knuth relation of the second kind transforms this word into

$$u_1 \cdots u_{j-2} u_j u_{j-1} v_k u_{j+1} \cdots u_{k-1}.$$

If  $j > 2$ , we now have  $u_{j-2} \leq u_{j-1} < u_j$ , so we can interchange  $u_j$  with  $u_{j-2}$ . We can continue in this way until  $u_j$  reaches the left end of the word. We have now transformed  $\text{rw}(T')v_k$  into  $\text{rw}(T)$  by elementary Knuth equivalences, so  $\text{rw}(T')v_k \equiv_K \text{rw}(T)$ .

For the induction step, assume  $\ell > 1$ . Let  $T''$  be the tableau  $T'$  with its first (longest) row erased. Then  $\text{rw}(T') = \text{rw}(T'')u_1 \cdots u_p$  where  $u_1 \leq \cdots \leq u_p$  is the weakly increasing sequence in the first row of  $T'$ . If  $u_p \leq v_k$ , then  $\text{rw}(T')v_k = \text{rw}(T)$ . Otherwise, assume  $v_k$  bumps  $u_j$  in the insertion  $T' \leftarrow v_k$ . By the result in the last paragraph,

$$\text{rw}(T')v_k \equiv_K \text{rw}(T'')u_j u_1 \cdots u_{j-1} v_k u_{j+1} \cdots u_p.$$

Now, by the induction hypothesis,  $\text{rw}(T'')u_j \equiv_K \text{rw}(T'' \leftarrow u_j)$ . Thus,

$$\text{rw}(T')v_k \equiv_K \text{rw}(T'' \leftarrow u_j)u'$$

where  $u'$  is  $u_1 \cdots u_p$  with  $u_j$  replaced by  $v_k$ . But, by definition of tableau insertion,  $\text{rw}(T'' \leftarrow u_j)u'$  is precisely  $\text{rw}(T)$ . This completes the induction step.  $\square$

**12.62. Example.** Let us illustrate how elementary Knuth equivalences implement the steps in the insertion  $T \leftarrow 3$ , where

$$T = \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 3 & 4 & 4 & 6 \\ \hline 2 & 2 & 4 & 5 & & \\ \hline 3 & 4 & & & & \\ \hline \end{array}$$

Appending a 3 at the right end of  $\text{rw}(T)$ , we first compute

$$34\ 2245\ 113446\ 3 \equiv_K 34\ 2245\ 1134436 \equiv_K 34\ 2245\ 1134346 \equiv_K$$

$$34\ 2245\ 1143346 \equiv_K 34\ 2245\ 1413346 \equiv_K 34\ 2245\ 4\ 113346.$$

The steps so far correspond to the insertion of 3 into the first row of  $T$ , which bumps the leftmost 4 into the second row. Continuing,

$$34\ 22454\ 113346 \equiv_K 34\ 22544\ 113346 \equiv_K 34\ 25244\ 113346 \equiv_K 34\ 5\ 2244\ 113346,$$

and now the incoming 4 has bumped the 5 into the third row. The process stops here with the word

$$3452244113346 = \text{rw} \left( \begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 3 & 3 & 4 & 6 \\ \hline 2 & 2 & 4 & 4 & & \\ \hline 3 & 4 & 5 & & & \\ \hline \end{array} \right) = \text{rw}(T \leftarrow 3).$$

This illustrates that  $\text{rw}(T)3 \equiv_K \text{rw}(T \leftarrow 3)$ .

**12.63. Definition: Increasing and Decreasing Subsequences.** Let  $w = w_1 w_2 \cdots w_n \in X^*$ . An *increasing subsequence* of  $w$  of length  $\ell$  is a subset  $I = \{i_1 < i_2 < \cdots < i_\ell\}$  of  $\{1, 2, \dots, n\}$  such that  $w_{i_1} \leq w_{i_2} \leq \cdots \leq w_{i_\ell}$ . A *decreasing subsequence* of  $w$  of length  $\ell$  is a subset  $I = \{i_1 < i_2 < \cdots < i_\ell\}$  such that  $w_{i_1} > w_{i_2} > \cdots > w_{i_\ell}$ . A *set of  $k$  disjoint increasing subsequences* of  $w$  is a set  $\{I_1, \dots, I_k\}$  of pairwise disjoint increasing subsequences of  $w$ . For each  $k \geq 1$ , let  $\text{inc}_k(w)$  be the maximum value of  $|I_1| + \cdots + |I_k|$  over all such sets. Similarly, let  $\text{dec}_k(w)$  be the maximum total length of a set of  $k$  disjoint decreasing subsequences of  $w$ .



**12.64. Theorem: Knuth Equivalence and Monotone Subsequences.** For all  $v, w \in X^*$  and all  $k \geq 1$ ,  $v \equiv_K w$  implies  $\text{inc}_k(v) = \text{inc}_k(w)$  and  $\text{dec}_k(v) = \text{dec}_k(w)$ .

*Proof.* It suffices to consider the case where  $v$  and  $w$  differ by a single elementary Knuth relation. First suppose

$$v = \mathbf{ayxz}\mathbf{b}, \quad w = \mathbf{ayzx}\mathbf{b}, \quad (x < y \leq z)$$

where the  $y$  occurs at position  $i$ . If  $I$  is an increasing subsequence of  $w$ , then  $i + 1$  and  $i + 2$  do not both belong to  $I$  (since  $z > x$ ). Therefore, if  $\{I_1, \dots, I_k\}$  is any set of  $k$  disjoint increasing subsequences of  $w$ , we can obtain a set  $\{I'_1, \dots, I'_k\}$  of disjoint increasing subsequences of  $v$  by replacing  $i + 1$  by  $i + 2$  and  $i + 2$  by  $i + 1$  in any  $I_j$  in which one of these indices appears. This implies that  $\text{inc}_k(w) \leq \text{inc}_k(v)$ .

To establish the opposite inequality, let  $\mathbf{I} = \{I_1, I_2, \dots, I_k\}$  be any set of  $k$  disjoint increasing subsequences of  $v$ . We will construct a set of  $k$  disjoint increasing subsequences of  $w$  having the same total size as  $\mathbf{I}$ . The device used in the previous paragraph works here, unless some member of  $\mathbf{I}$  (say  $I_1$ ) contains both  $i + 1$  and  $i + 2$ . In this case, we cannot have  $i \in I_1$ , since  $y > x$ . If no other member of  $\mathbf{I}$  contains  $i$ , we replace  $I_1$  by  $(I_1 \sim \{i + 2\}) \cup \{i\}$ , which is an increasing subsequence of  $w$ . On the other hand, suppose  $i + 1, i + 2 \in I_1$ , and some other member of  $\mathbf{I}$  (say  $I_2$ ) contains  $i$ . Write

$$\begin{aligned} I_1 &= \{j_1 < j_2 < \dots < j_r < i + 1 < i + 2 < j_{r+1} < \dots < j_p\}, \\ I_2 &= \{k_1 < k_2 < \dots < k_s < i < k_{s+1} < \dots < k_q\}, \end{aligned}$$

and note that  $v_{j_r} \leq x < z \leq v_{j_{r+1}}$  and  $v_{k_s} \leq y \leq v_{k_{s+1}}$ . Replace these two disjoint increasing subsequences of  $v$  by

$$\begin{aligned} I'_1 &= \{j_1 < j_2 < \dots < j_r < i + 2 < k_{s+1} < \dots < k_q\}, \\ I'_2 &= \{k_1 < k_2 < \dots < k_s < i < i + 1 < j_{r+1} < \dots < j_p\}. \end{aligned}$$

Since  $w_{j_r} \leq x \leq w_{k_{s+1}}$  and  $w_{k_s} \leq y \leq z \leq w_{j_{r+1}}$ ,  $I'_1$  and  $I'_2$  are two disjoint increasing subsequences of  $w$  having the same total length as  $I_1$  and  $I_2$ . This completes the proof that  $\text{inc}_k(w) \geq \text{inc}_k(v)$ .

Similar reasoning (left as an exercise for the reader) proves the result in the case where

$$v = \mathbf{axyz}\mathbf{b}, \quad w = \mathbf{axzy}\mathbf{b}, \quad (x \leq y < z).$$

We also let the reader prove the statement about decreasing subsequences. □

**12.65. Theorem: Subsequences and the Shape of Insertion Tableaux.** Let  $w \in X^*$  and suppose  $P(w)$  has shape  $\lambda$ . For all  $k \geq 1$ ,

$$\text{inc}_k(w) = \lambda_1 + \dots + \lambda_k, \quad \text{dec}_k(w) = \lambda'_1 + \dots + \lambda'_k.$$

In particular,  $\lambda_1$  is the length of the longest increasing subsequence of  $w$ , whereas  $\ell(\lambda)$  is the length of the longest decreasing subsequence of  $w$ .

*Proof.* Let  $w' = \text{rw}(P(w))$ . We know  $w \equiv_K w'$  by 12.61, so  $\text{inc}_k(w) = \text{inc}_k(w')$  and  $\text{dec}_k(w) = \text{dec}_k(w')$  by 12.64. So we need only prove

$$\text{inc}_k(w') = \lambda_1 + \dots + \lambda_k, \quad \text{dec}_k(w') = \lambda'_1 + \dots + \lambda'_k.$$

Now,  $w'$  consists of increasing sequences of letters of successive lengths  $\lambda_l, \dots, \lambda_2, \lambda_1$  (where  $l = \ell(\lambda)$ ). By taking  $I_1, I_2, \dots, I_k$  to be the set of positions of the last  $k$  of these sequences, we

obtain  $k$  disjoint increasing subsequences of  $w'$  of length  $\lambda_1 + \cdots + \lambda_k$ . Therefore,  $\text{inc}_k(w') \geq \lambda_1 + \cdots + \lambda_k$ .

On the other hand, let  $\{I_1, \dots, I_k\}$  be any  $k$  disjoint increasing subsequences of  $w'$ . Each position  $i$  in  $w'$  is associated to a particular box in the diagram of  $\lambda$ , via 12.56. For example, position 1 corresponds to the first box in the last row, while the last position corresponds to the last box in the first row. For each position  $i$  that belongs to some  $I_j$ , place an X in the corresponding box in the diagram of  $\lambda$ . Since entries in a given column of  $P(w)$  strictly decrease reading from bottom to top, the X's coming from a given increasing subsequence  $I_j$  must all lie in different columns of the diagram. It follows that every column of the diagram contains  $k$  or fewer X's. Suppose we push all these X's up their columns as far as possible. Then all the X's in the resulting figure must lie in the top  $k$  rows of  $\lambda$ . It follows that the number of X's, which is  $|I_1| + \cdots + |I_k|$ , cannot exceed  $\lambda_1 + \cdots + \lambda_k$ . This gives  $\text{inc}_k(w') \leq \lambda_1 + \cdots + \lambda_k$ . The proof for  $\text{dec}_k(w)$  is similar, and is left as an exercise.  $\square$

**12.66. Theorem: Knuth Equivalence vs. Tableau Shape.** For all  $v, w \in X^*$ ,  $v \equiv_K w$  implies that  $P(v)$  and  $P(w)$  have the same shape.

*Proof.* Let  $\lambda$  and  $\mu$  be the shapes of  $P(v)$  and  $P(w)$ , respectively. Using 12.64 and 12.65, we see that  $v \equiv_K w$  implies

$$\lambda_k = \text{inc}_k(v) - \text{inc}_{k-1}(v) = \text{inc}_k(w) - \text{inc}_{k-1}(w) = \mu_k \quad (k \geq 1). \quad \square$$

**12.67. Example.** Consider the word  $w = 35164872$ . As shown in Figure 10.1, we have

$$P(w) = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 6 & 7 \\ \hline 3 & 4 & 8 & \\ \hline 5 & & & \\ \hline \end{array}$$

Since the shape is  $\lambda = (4, 3, 1)$ , the longest increasing subsequence of  $w$  has length 4. Two such subsequences are  $I_1 = \{1, 2, 4, 7\}$  (corresponding to the subword 3567) and  $I_2 = \{1, 2, 4, 6\}$ . Note that the first row of  $P(w)$ , namely 1267, does *not* appear as a subword of  $w$ . Since the column lengths of  $\lambda$  are  $(3, 2, 2, 1)$ , the longest length of two disjoint decreasing subsequences of  $w$  is  $3 + 2 = 5$ . For example, we could take  $I_1 = \{6, 7, 8\}$  and  $I_2 = \{4, 5\}$  to achieve this. Note that  $w' = \text{rw}(P(w)) = 5\ 348\ 1267$ . To illustrate the end of the previous proof, consider the two disjoint increasing subsequences  $I_1 = \{1, 4\}$  and  $I_2 = \{2, 3, 7, 8\}$  of  $w'$  (this pair does not achieve the maximum length for such subsequences). Drawing X's in the boxes of the diagram associated to the positions in  $I_1$  (resp.  $I_2$ ) produces

$$\begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & X & \\ \hline X & & & \\ \hline \end{array} \quad \left( \text{resp.} \quad \begin{array}{|c|c|c|c|} \hline & & X & X \\ \hline X & X & & \\ \hline & & & \\ \hline \end{array} \right).$$

Combining these diagrams and pushing the X's up as far as they will go, we get

$$\begin{array}{|c|c|c|c|} \hline X & X & X & X \\ \hline X & & X & \\ \hline & & & \\ \hline \end{array}$$

So, indeed, the combined length of  $I_1$  and  $I_2$  does not exceed  $\lambda_1 + \lambda_2$ .

The next lemma provides the remaining ingredients needed to establish that  $P$ -equivalence and Knuth equivalence are the same.

**12.68. Lemma.** Suppose  $v, w \in X^*$  and  $z$  is the largest letter appearing in both  $v$  and  $w$ . Let  $v'$  (resp.  $w'$ ) be the word obtained by erasing the rightmost  $z$  from  $v$  (resp.  $w$ ). If  $v \equiv_K w$ , then  $v' \equiv_K w'$ . Furthermore, if  $T = P(v)$  and  $T' = P(v')$ , then  $T'$  can be obtained from  $T$  by erasing the rightmost box containing  $z$ .

*Proof.* Write  $v = azb$  and  $w = czd$  where  $a, b, c, d \in X^*$  and  $z$  does not appear in  $b$  or  $d$ . First assume that  $v$  and  $w$  differ by a single elementary Knuth relation. If the triple of letters affected by this relation are part of the subword  $a$ , then  $a \equiv_K c$  and  $b = d$ , so  $v' = ab \equiv_K cd = w'$ . Similarly, the result holds if the triple of letters is part of the subword  $b$ . The next possibility is that

$$v = a'yxzb, \quad w = a'yzxb \quad (x < y \leq z)$$

(or vice versa). Then  $v' = a'ymb = w'$ , so certainly  $v' \equiv_K w'$ . Another possibility is that

$$v = a'xzyb', \quad w = a'zxyb' \quad (x \leq y < z)$$

(or vice versa), and again  $v' = a'xyb' = w'$ . Since the  $z$  under consideration is the rightmost occurrence of the largest letter in both  $v$  and  $w$ , the possibilities already considered are the only elementary Knuth relations that involve this symbol. So the result holds when  $v$  and  $w$  differ by one elementary Knuth relation. Now, if  $v = v^0, v^1, v^2, \dots, v^k = w$  is a sequence of words as in 12.59, we can write each  $v^i = a^i z b^i$  where  $z$  does not appear in  $b^i$ . Letting  $(v^i)' = a^i b^i$  for each  $i$ , the chain  $v' = (v^0)', (v^1)', \dots, (v^k)' = w'$  proves that  $v' \equiv_K w'$ .

Now consider the actions of the tableau insertion algorithm applied to  $v = azb$  and to  $v' = ab$ . We prove the statement about  $T$  and  $T'$  by induction on the length of  $b$ . The statement holds if  $b$  is empty. Assume  $b$  has length  $k > 0$  and the statement is known for smaller values of  $k$ . Write  $b = b'x$  where  $x \in X$ . Then  $T'_1 = P(ab')$  is the tableau  $T_1 = P(azb')$  with the rightmost  $z$  erased. By definition,  $T' = T'_1 \leftarrow x$  and  $T = T_1 \leftarrow x$ . When we insert the  $x$  into these two tableaux, the bumping paths will be the same (and hence the desired result holds), unless  $x$  bumps the rightmost  $z$  in  $T_1$ . If this happens, the rightmost  $z$  (which must have been the only  $z$  in its row) will get bumped into the next lower row. It will come to rest there without bumping anything else, and it will still be the rightmost  $z$  in the tableau. Thus it is still true that erasing this  $z$  in  $T$  produces  $T'$ . The induction is therefore complete.  $\square$

**12.69. Theorem:  $P$ -Equivalence vs. Knuth Equivalence.** For all  $v, w \in X^*$ ,  $v \equiv_P w$  iff  $v \equiv_K w$ .

*Proof.* First, if  $v \equiv_P w$ , then 12.61 shows that  $v \equiv_K \text{rw}(P(v)) = \text{rw}(P(w)) \equiv_K w$ , so  $v \equiv_K w$  by transitivity of  $\equiv_K$ . Conversely, assume  $v \equiv_K w$ . We prove  $v \equiv_P w$  by induction on the length  $k$  of  $v$ . For  $k \leq 1$ , we have  $v = w$  and so  $v \equiv_P w$ . Now assume  $k > 1$  and the result is known for words of length  $k - 1$ . Write  $v = azb$  and  $w = czd$  where  $z$  is the largest symbol in  $v$  and  $w$  and  $z$  does not occur in  $b$  or  $d$ . Write  $v' = ab$  and  $w' = cd$ . By 12.68,  $v' \equiv_K w'$ ,  $P(v')$  is  $P(v)$  with the rightmost  $z$  erased, and  $P(w')$  is  $P(w)$  with the rightmost  $z$  erased. By induction,  $P(v') = P(w')$ . If we knew that  $P(v)$  and  $P(w)$  had the same shape, it would follow that  $P(v) = P(w)$ . But  $P(v)$  and  $P(w)$  do have the same shape, thanks to 12.66. So  $v \equiv_P w$ .  $\square$

We conclude with an application of 12.65.

**12.70. Erdős-Szekeres Subsequence Theorem.** Every word of length exceeding  $mn$  either has an increasing subsequence of length  $m + 1$  or a decreasing subsequence of length  $n + 1$ .

*Proof.* Suppose  $w$  is a word with no increasing subsequence of length  $m + 1$  and no decreasing subsequence of length  $n + 1$ . Let  $\lambda$  be the shape of  $P(w)$ . Then 12.65 implies that  $\lambda_1 \leq m$  and  $\ell(\lambda) \leq n$ . Therefore the length of  $w$ , which is  $|\lambda|$ , can be no greater than  $\lambda_1 \ell(\lambda) \leq mn$ .  $\square$

## 12.12 Pfaffians and Perfect Matchings

Given a *square* matrix  $A$  with  $N$  rows and  $N$  columns, we have defined the *determinant* of  $A$  by the formula

$$\det(A) = \sum_{w \in S_N} \operatorname{sgn}(w) \prod_{i=1}^N A(i, w(i)).$$

This section studies *Pfaffians*, which are numbers associated to a *triangular* array of numbers  $(a_{i,j} : 1 \leq i < j \leq N)$  where  $N$  is even. Pfaffians arise in the theory of skew-symmetric matrices.

**12.71. Definition: Skew-Symmetric Matrices.** An  $N \times N$  matrix  $A$  is called *skew-symmetric* iff  $A^t = -A$  iff  $A(i, j) = -A(j, i)$  for  $1 \leq i, j \leq N$ .

If  $A$  is a real or complex skew-symmetric matrix, then  $A(i, i) = 0$  for all  $i$ . Moreover,  $A$  is completely determined by the triangular array of numbers  $(A(i, j) : 1 \leq i < j \leq N)$  lying strictly above the main diagonal. The starting point for the theory of Pfaffians is the observation that, for  $N$  even and  $A$  skew-symmetric,  $\det(A)$  is always a perfect square. (For  $N$  odd, the condition  $A^t = -A$  can be used to show that  $\det(A) = 0$ .)

**12.72. Example.** A general skew-symmetric  $2 \times 2$  matrix has the form  $A = \begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}$ .

In this case,  $\det(A) = a^2$  is a square. A skew-symmetric  $4 \times 4$  matrix looks like

$$A = \begin{bmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{bmatrix}.$$

A somewhat tedious calculation reveals that

$$\begin{aligned} \det(A) &= a^2 f^2 + b^2 e^2 + c^2 d^2 - 2abef + 2acdf - 2bcde \\ &= (af + cd - be)^2. \end{aligned}$$

The remainder of this section develops the theory needed to explain the phenomenon observed in the last example.

**12.73. Definition: Pfaffians.** Suppose  $N$  is even and  $A$  is a skew-symmetric  $N \times N$  matrix. Let  $\operatorname{SPf}_N$  be the set of all permutations  $w \in S_N$  such that

$$w_1 < w_3 < w_5 < \cdots < w_{N-1}, \quad w_1 < w_2, \quad w_3 < w_4, \quad w_5 < w_6, \quad \dots, \quad w_{N-1} < w_N.$$

The *Pfaffian* of  $A$ , denoted  $\operatorname{Pf}(A)$ , is the number

$$\operatorname{Pf}(A) = \sum_{w \in \operatorname{SPf}_N} \operatorname{sgn}(w) A(w_1, w_2) A(w_3, w_4) A(w_5, w_6) \cdots A(w_{N-1}, w_N).$$

**12.74. Example.** If  $N = 2$ ,  $\operatorname{SPf}_2 = \{12\}$  and  $\operatorname{Pf}(A) = A(1, 2)$  (we write permutations in one-line form here). If  $N = 4$ ,  $\operatorname{SPf}_4 = \{1234, 1423, 1324\}$  and

$$\operatorname{Pf}(A) = A(1, 2)A(3, 4) + A(1, 4)A(2, 3) - A(1, 3)A(2, 4).$$

For a general  $N \times N$  matrix  $A$ ,  $\det(A)$  is a sum of  $|S_N| = N!$  terms. Similarly, for a skew-symmetric matrix  $A$ ,  $\text{Pf}(A)$  is a sum of  $|\text{SPf}_N|$  terms.

**12.75. Theorem: Size of  $\text{SPf}_N$ .** For each even  $N$ ,  $|\text{SPf}_N| = 1 \times 3 \times 5 \times \cdots \times (N-1)$ .

*Proof.* We can construct each permutation  $w \in \text{SPf}_N$  as follows. First,  $w_1$  must be 1. There are  $N-1$  choices for  $w_2$ , which can be anything other than 1. To finish building  $w$ , choose an arbitrary permutation  $v = v_1 v_2 \cdots v_{N-2} \in \text{SPf}_{N-2}$ . For  $1 \leq i \leq N-2$ , set

$$w_{i+2} = \begin{cases} v_i + 1 & \text{if } v_i < w_2 - 1 \\ v_i + 2 & \text{otherwise.} \end{cases}$$

Informally, we are renumbering the  $v$ 's to use symbols in  $\{1, 2, \dots, N\} \sim \{w_1 = 1, w_2\}$  and then appending this word to  $w_1 w_2$ . By the product rule,  $|\text{SPf}_N| = (N-1) \times |\text{SPf}_{N-2}|$ . Since  $|\text{SPf}_2| = 1$ , the formula in the theorem follows by induction.  $\square$

Recall that the Laplace expansions in 9.48 provide recursive formulas for evaluating determinants. Similar recursive formulas exist for evaluating Pfaffians. The key difference is that *two* rows and columns get erased at each stage, whereas in Laplace expansions only one row and column get erased at a time.

**12.76. Theorem: Pfaffian Expansion along Row 1.** Suppose  $N$  is even and  $A$  is an  $N \times N$  skew-symmetric matrix. For each  $i < j$ , let  $A[[i, j]]$  be the matrix obtained from  $A$  by deleting row  $i$  and row  $j$  and column  $i$  and column  $j$ ; this is a skew-symmetric matrix of size  $(N-2) \times (N-2)$ . We have

$$\text{Pf}(A) = \sum_{j=2}^N (-1)^j A(1, j) \text{Pf}(A[[1, j]]).$$

*Proof.* By definition,

$$\text{Pf}(A) = \sum_{w \in \text{SPf}_N} \text{sgn}(w) \prod_{\substack{i=1 \\ i \text{ odd}}}^N A(w_i, w_{i+1}).$$

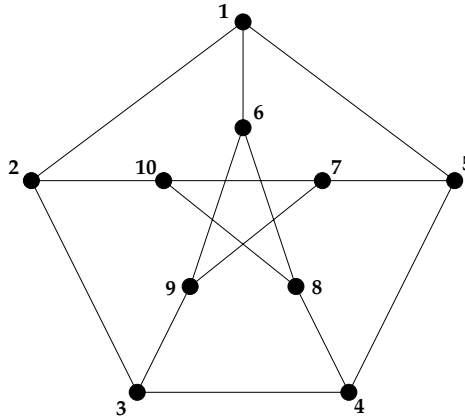
By the proof of 12.75, there is a bijection  $\text{SPf}_N \rightarrow \{2, 3, \dots, N\} \times \text{SPf}_{N-2}$  that maps  $w \in \text{SPf}_N$  to  $(j, v)$ , where  $j = w_2$  and  $v$  is obtained from  $w_3 w_4 \cdots w_N$  by renumbering the symbols to be  $1, 2, \dots, N-2$ . We will use this bijection to change the indexing set for the summation from  $\text{SPf}_N$  to  $\{2, \dots, N\} \times \text{SPf}_{N-2}$ . Counting inversions, we see that  $\text{inv}(w) = \text{inv}(v) + j - 2$  since  $w_2 = j$  exceeds  $j-2$  symbols to its right. So  $\text{sgn}(w) = (-1)^j \text{sgn}(v)$ . Next,  $A(w_1, w_2) = A(1, j)$ . For odd  $i > 1$ , it follows from the definitions that  $A(w_i, w_{i+1}) = A[[1, j]](v_{i-2}, v_{i-1})$ . Putting all this information into the formula, we see that

$$\text{Pf}(A) = \sum_{j=2}^N (-1)^j A(1, j) \sum_{v \in \text{SPf}_{N-2}} \text{sgn}(v) \prod_{\substack{i=1 \\ i \text{ odd}}}^{N-2} A[[1, j]](v_i, v_{i+1}).$$

The inner sum is precisely  $\text{Pf}(A[[1, j]])$ , so the proof is complete.  $\square$

**12.77. Example.** Let us compute the Pfaffian of the matrix

$$A = \begin{bmatrix} 0 & x & -y & 0 & 0 & 0 \\ -x & 0 & 0 & y & 0 & 0 \\ y & 0 & 0 & x & -y & 0 \\ 0 & -y & -x & 0 & 0 & y \\ 0 & 0 & y & 0 & 0 & x \\ 0 & 0 & 0 & -y & -x & 0 \end{bmatrix}.$$



**FIGURE 12.14**

Graph used to illustrate perfect matchings.

Expanding along row 1 gives

$$\text{Pf}(A) = x \text{Pf} \begin{bmatrix} 0 & x & -y & 0 \\ -x & 0 & 0 & y \\ y & 0 & 0 & x \\ 0 & -y & -x & 0 \end{bmatrix} - (-y) \text{Pf} \begin{bmatrix} 0 & y & 0 & 0 \\ -y & 0 & 0 & y \\ 0 & 0 & 0 & x \\ 0 & -y & -x & 0 \end{bmatrix}.$$

By expanding these  $4 \times 4$  Pfaffians in the same way, or by using the formula in 12.74, we obtain

$$\text{Pf}(A) = x(x^2 + y^2) + y(xy) = x^3 + 2xy^2.$$

The combinatorial significance of this Pfaffian evaluation will be revealed in §12.13.

Pfaffians are closely related to perfect matchings of graphs, which we now discuss.

**12.78. Definition: Perfect Matchings.** Let  $G$  be a simple graph with vertex set  $V$  and edge set  $E$ . A *perfect matching* of  $G$  is a subset  $M$  of  $E$  such that each  $v \in V$  is the endpoint of exactly one edge in  $M$ . Let  $\text{PM}(G)$  be the set of perfect matchings of  $G$ .

**12.79. Example.** For the graph shown in Figure 12.14, one perfect matching is

$$M_1 = \{\{1, 6\}, \{2, 10\}, \{3, 9\}, \{4, 8\}, \{5, 7\}\}.$$

Another perfect matching is

$$M_2 = \{\{1, 2\}, \{3, 4\}, \{5, 7\}, \{6, 9\}, \{8, 10\}\}.$$

A perfect matching on a graph  $G$  is a set partition of the vertex set of  $G$  into blocks of size 2 where each such block is an edge of  $G$ . Therefore, if  $G$  has  $N$  vertices and a perfect matching exists for  $G$ , then  $N$  must be even. The next result shows that perfect matchings on a *complete* graph can be encoded by permutations in  $\text{SPf}_N$ .

**12.80. Theorem: Perfect Matchings on a Complete Graph.** Suppose  $N$  is even and  $K_N$  is the simple graph with vertex set  $\{1, 2, \dots, N\}$  and edge set  $\{\{i, j\} : 1 \leq i < j \leq N\}$ . The map  $f : \text{SPf}_N \rightarrow \text{PM}(K_N)$  defined by

$$f(w_1 w_2 \cdots w_N) = \{\{w_1, w_2\}, \{w_3, w_4\}, \dots, \{w_{N-1}, w_N\}\}$$

is a bijection. Consequently,

$$|\text{PM}(K_N)| = 1 \times 3 \times 5 \times \cdots \times (N-1).$$

*Proof.* Note first that  $f$  does map into the set  $\text{PM}(K_N)$ . Next, a matching  $M \in \text{PM}(K_N)$  is a set of  $N/2$  edges  $M = \{\{i_1, i_2\}, \{i_3, i_4\}, \dots, \{i_{N-1}, i_N\}\}$ . Since  $\{i, j\} = \{j, i\}$ , we can choose the notation so that  $i_1 < i_2, i_3 < i_4, \dots$ , and  $i_{N-1} < i_N$ . Similarly, since the  $N/2$  edges of  $M$  can be presented in any order, we can change notation again (if needed) to arrange that  $i_1 < i_3 < i_5 < \cdots < i_{N-1}$ . Then the permutation  $w = i_1 i_2 i_3 \cdots i_N \in \text{SPf}_N$  satisfies  $f(w) = M$ . Thus  $f$  maps *onto*  $\text{PM}(K_N)$ . To see that  $f$  is one-to-one, suppose  $v = j_1 j_2 j_3 \cdots j_N$  is another element of  $\text{SPf}_N$  such that  $f(v) = M = f(w)$ . We must have  $j_1 = 1 = i_1$ . Since  $M$  has only one edge incident to vertex 1, and since  $\{i_1, i_2\} \in M$  and  $\{j_1, j_2\} \in M$  by definition of  $f$ , we conclude that  $i_2 = j_2$ . Now  $i_3$  and  $j_3$  must both be the smallest vertex in the set  $\{1, 2, \dots, N\} \sim \{i_1, i_2\}$ , so  $i_3 = j_3$ . Then  $i_4 = j_4$  follows, as above, since  $M$  is a perfect matching. Continuing similarly, we see that  $i_k = j_k$  for all  $k$ , so  $v = w$  and  $f$  is one-to-one. Since  $f$  is a bijection, the formula for  $|\text{PM}(K_N)|$  follows from 12.75.  $\square$

The preceding theorem leads to the following combinatorial interpretation for Pfaffians. Given a perfect matching  $M \in \text{PM}(K_N)$ , use 12.80 to write  $M = f(w)$  for some  $w \in \text{SPf}_N$ . Define the *sign* of  $M$  to be  $\text{sgn}(w)$ , and define the *weight* of  $M$  to be

$$\text{wt}(M) = \prod_{\{i,j\} \in M} x_{i,j} = \prod_{\substack{i=1 \\ i \text{ odd}}}^N x_{w_i, w_{i+1}},$$

where the  $x_{i,j}$  (for  $1 \leq i < j \leq N$ ) are indeterminates. Let  $X$  be the skew-symmetric matrix with entries  $x_{i,j}$  above the main diagonal. It follows from 12.80 and the definition of a Pfaffian that

$$\sum_{M \in \text{PM}(K_N)} \text{sgn}(M) \text{wt}(M) = \text{Pf}(X).$$

More generally, we have the following result.

**12.81. Theorem: Pfaffians and Perfect Matchings.** Let  $N$  be even, and let  $G$  be a simple graph with vertex set  $V = \{1, 2, \dots, N\}$  and edge set  $E(G)$ . Let  $X = X(G)$  be the skew-symmetric matrix with entries

$$X(i, j) = \begin{cases} x_{i,j} & \text{if } i < j \text{ and } \{i, j\} \in E(G) \\ -x_{i,j} & \text{if } i > j \text{ and } \{i, j\} \in E(G) \\ 0 & \text{otherwise.} \end{cases}$$

Then  $\sum_{M \in \text{PM}(G)} \text{sgn}(M) \text{wt}(M) = \text{Pf}(X(G))$ .

*Proof.* We have already observed that

$$\sum_{M \in \text{PM}(K_N)} \text{sgn}(M) \text{wt}(M) = \text{Pf}(X(K_N)). \quad (12.11)$$

Given the graph  $G$ , let  $\epsilon$  be the evaluation homomorphism (see 7.102) that sends  $x_{i,j}$  to  $x_{i,j}$  if  $\{i, j\} \in E(G)$ , and sends  $x_{i,j}$  to 0 if  $\{i, j\} \notin E(G)$ . Applying  $\epsilon$  to the left side of (12.11) produces

$$\sum_{M \in \text{PM}(G)} \text{sgn}(M) \text{wt}(M),$$

since all matchings of  $K_N$  that use an edge not in  $E(G)$  are mapped to zero. On the other hand, since  $\epsilon$  is a ring homomorphism and the Pfaffian of a matrix is a polynomial in the entries of the matrix, we can compute  $\epsilon(\text{Pf}(X(K_N)))$  by applying  $\epsilon$  to each entry of  $X(K_N)$  and taking the Pfaffian of the resulting matrix. So, applying  $\epsilon$  to the right side of (12.11) gives

$$\epsilon(\text{Pf}(X(K_N))) = \text{Pf}(\epsilon(X(K_N))) = \text{Pf}(X(G)). \quad \square$$

**12.82. Remark.** The last result shows that  $\text{Pf}(X(G))$  is a *signed* sum of distinct monomials, where there is one monomial for each perfect matching of  $G$ . Because of the signs, one cannot compute  $|\text{PM}(G)|$  by setting  $x_{i,j} = 1$  for each  $\{i, j\} \in E(G)$ . However, for certain graphs  $G$ , one can introduce extra signs into the upper part of the matrix  $X(G)$  to counteract the sign arising from  $\text{sgn}(M)$ . This process is illustrated in the next section.

We can now give a combinatorial proof of the main result linking Pfaffians and determinants.

**12.83. Theorem: Pfaffians vs. Determinants.** For every even  $N$  and every  $N \times N$  skew-symmetric matrix  $A$ ,  $\det(A) = \text{Pf}(A)^2$ .

*Proof.* First we use the skew-symmetry of  $A$  to cancel some terms in the sum

$$\det(A) = \sum_{w \in S_N} \text{sgn}(w) \prod_{i=1}^N A(i, w(i)).$$

We will cancel every term indexed by a permutation  $w$  whose functional digraph contains at least one cycle of odd length (cf. §3.6). If  $w$  has a cycle of length 1, then  $w(i) = i$  for some  $i$ . So  $A(i, w(i)) = A(i, i) = 0$  by skew-symmetry, and the term indexed by this  $w$  is zero. On the other hand, suppose  $w$  has no fixed points, but  $w$  does have at least one cycle of odd length. Among all the odd-length cycles of  $w$ , choose the cycle  $(i_1, i_2, \dots, i_k)$  whose minimum element is as small as possible. Reverse the orientation of this cycle to get a permutation  $w' \neq w$ . For example, if  $w = (3, 8, 4)(2, 5, 7)(1, 6)(9, 10)$ , then  $w' = (3, 8, 4)(7, 5, 2)(1, 6)(9, 10)$ . In general,  $\text{sgn}(w') = \text{sgn}(w)$  since  $w$  and  $w'$  have the same cycle structure (see 9.34). However, since  $k$  is odd and  $A$  is skew-symmetric,

$$A(i_1, i_2)A(i_2, i_3) \cdots A(i_{k-1}, i_k)A(i_k, i_1) = -A(i_2, i_1)A(i_3, i_2) \cdots A(i_k, i_{k-1})A(i_1, i_k).$$

It follows that the term in  $\det(A)$  indexed by  $w'$  is the negative of the term in  $\det(A)$  indexed by  $w$ , so this pair of terms cancels. Since  $w \mapsto w'$  is an involution, we conclude that

$$\det(A) = \sum_{w \in S_N^{ev}} \text{sgn}(w) \prod_{i=1}^N A(i, w(i)),$$

where  $S_N^{ev}$  denotes the set of permutations of  $N$  objects with only even-length cycles.

The next step is to compare the terms in this sum to the terms in  $\text{Pf}(A)^2$ . Using the distributive law to square the defining formula for  $\text{Pf}(A)$ , we see that

$$\text{Pf}(A)^2 = \sum_{u \in \text{SPf}_N} \sum_{v \in \text{SPf}_N} \text{sgn}(u) \text{sgn}(v) \prod_{i \text{ odd}} [A(u_i, u_{i+1})A(v_i, v_{i+1})].$$

Given  $w \in S_N^{ev}$  indexing an uncanceled term in  $\det(A)$ , we associate a pair  $(u, v) \in \text{SPf}_N^2$  indexing a summand in  $\text{Pf}(A)^2$  as follows. Consider the functional digraph  $G(w)$  with vertex set  $\{1, 2, \dots, N\}$  and edge set  $\{(i, w(i)) : 1 \leq i \leq N\}$ , which is a disjoint union of cycles. Define a perfect matching  $M_1$  on  $G(w)$  (viewed as an undirected graph) by starting at the



minimum element in each cycle and including every other edge as one travels around the cycle. Define another perfect matching  $M_2$  on  $G(w)$  by taking all the edges not used in  $M_1$ . Finally, let  $u$  and  $v$  be the permutations in  $\text{SPf}_N$  that encode  $M_1$  and  $M_2$  via the bijection in 12.80. For example, if  $w = (1, 5, 2, 8, 6, 3)(4, 7)$ , then  $M_1 = \{\{1, 5\}, \{2, 8\}, \{6, 3\}, \{4, 7\}\}$  and  $M_2 = \{\{5, 2\}, \{8, 6\}, \{3, 1\}, \{7, 4\}\}$ , so  $u = 15283647$  and  $v = 13254768$ . The association  $w \mapsto (u, v)$  is a bijection from  $S_N^{ev}$  to  $\text{SPf}_N^2$ . To compute the inverse map, one need only take the union of the perfect matchings encoded by  $u$  and  $v$ . This produces a graph that is a disjoint union of cycles of even length, as is readily checked. One can restore the directions on each cycle by recalling that the outgoing edge from the minimum element in each cycle belongs to the matching encoded by  $u$ . For example, the pair  $(u', v') = (15234867, 12374856)$  maps to  $w' = (1, 5, 6, 7, 3, 2)(4, 8)$  under the inverse bijection.

Throughout the following discussion, assume  $w \in S_N^{ev}$  corresponds to  $(u, v) \in \text{SPf}_N^2$ . To complete the proof, it suffices to show that the term in  $\det(A)$  indexed by  $w$  equals the term in  $\text{Pf}(A)^2$  indexed by  $(u, v)$ . Write  $w$  in cycle form as

$$w = (m_1, n_1, \dots, z_1)(m_2, n_2, \dots, z_2) \cdots (m_k, n_k, \dots, z_k)$$

where  $m_1 < m_2 < \cdots < m_k$  are the minimum elements in their cycles. Define two words (permutations in one-line form)

$$\begin{aligned} u^* &= m_1 n_1 \cdots z_1 m_2 n_2 \cdots z_2 \cdots m_k n_k \cdots z_k; \\ v^* &= n_1 \cdots z_1 m_1 n_2 \cdots z_2 m_2 \cdots n_k \cdots z_k m_k. \end{aligned}$$

Thus  $u^*$  is obtained by erasing the parentheses in the particular cycle notation for  $w$  just mentioned, and  $v^*$  is obtained similarly after first cycling the values in each cycle one step to the left. Since each  $m_i$  is the smallest value in its cycle, it follows that

$$\text{inv}(v^*) = N - k + \text{inv}(u^*) \quad (k = \text{cyc}(w)).$$

Therefore  $\text{sgn}(u^*) \text{sgn}(v^*) = (-1)^{N - \text{cyc}(w)} = \text{sgn}(w)$ . Since all the edges  $(i, w(i))$  in  $G(w)$  arise by pairing off consecutive letters in  $u^*$  and  $v^*$ , we have

$$\text{sgn}(w) \prod_{i=1}^N A(i, w(i)) = \text{sgn}(u^*) \text{sgn}(v^*) \prod_{i \text{ odd}} [A(u_i^*, u_{i+1}^*) A(v_i^*, v_{i+1}^*)].$$

We now transform the right side to the term indexed by  $(u, v)$  in  $\text{Pf}(A)^2$ , as follows. Note that the words  $u^*$  and  $v^*$  provide *non-standard* encodings of the perfect matchings  $M_1$  and  $M_2$  encoded by  $u$  and  $v$  (the edges of the matchings are found by grouping pairs of consecutive symbols in  $u^*$  and  $v^*$ ). To get to the standard encodings, first reverse each pair of consecutive letters  $u_i^*, u_{i+1}^*$  in  $u^*$  such that  $u_i^* > u_{i+1}^*$  and  $i$  is odd. Each such reversal causes  $\text{sgn}(u^*)$  to change, but this change is balanced by the fact that  $A(u_{i+1}^*, u_i^*) = -A(u_i^*, u_{i+1}^*)$ . Similarly, we can reverse pairs of consecutive letters in  $v^*$  that are out of order. The next step is to sort the pairs in  $u^*$  to force  $u_1 < u_3 < u_5 < \cdots < u_{N-1}$ . This sorting can be achieved by repeatedly swapping adjacent pairs  $a < b; c < d$  in the word, where  $a > c$ . The swap  $abcd \mapsto cdab$  can be achieved by applying the two transpositions  $(a, c)$  and  $(b, d)$  on the left. So this modification of  $u^*$  does not change  $\text{sgn}(u^*)$ , nor does it affect the product of the factors  $A(u_i^*, u_{i+1}^*)$  (since multiplication is commutative). Similarly, we can sort the pairs in  $v^*$  to obtain  $v$  without changing the formula. We conclude finally that

$$\begin{aligned} \text{sgn}(w) \prod_{i=1}^N A(i, w(i)) &= \text{sgn}(u^*) \text{sgn}(v^*) \prod_{i \text{ odd}} [A(u_i^*, u_{i+1}^*) A(v_i^*, v_{i+1}^*)] \\ &= \text{sgn}(u) \text{sgn}(v) \prod_{i \text{ odd}} [A(u_i, u_{i+1}) A(v_i, v_{i+1})]. \quad \square \end{aligned}$$

The following example illustrates the calculations at the end of the preceding proof.

**12.84. Example.** Suppose  $w = (3, 8)(11, 4, 2, 9)(1, 10, 6, 7, 5, 12) \in S_{12}^{ev}$ , so  $k = \text{cyc}(w) = 3$ . We begin by writing the standard cycle notation for  $w$ :

$$w = (1, 10, 6, 7, 5, 12)(2, 9, 11, 4)(3, 8).$$

Next we set

$$u^* = 1, 10; 6, 7; 5, 12; 2, 9; 11, 4; 3, 8; \quad v^* = 10, 6; 7, 5; 12, 1; 9, 11; 4, 2; 8, 3.$$

Observe that  $\text{inv}(v^*) = \text{inv}(u^*) + (12 - 3)$  due to the cyclic shifting of 1, 2, 3, so that  $\text{sgn}(u^*)\text{sgn}(v^*) = (-1)^{12-3} = \text{sgn}(w)$ . Now we modify  $u^*$  and  $v^*$  so that the elements in each pair increase:

$$u' = 1, 10; 6, 7; 5, 12; 2, 9; 4, 11; 3, 8; \quad v' = 6, 10; 5, 7; 1, 12; 9, 11; 2, 4; 3, 8.$$

Note that  $\text{sgn}(u') = -\text{sgn}(u^*)$  since we switched 11 and 4, but this is offset by the fact that  $A(11, 4) = -A(4, 11)$ . So  $\text{sgn}(u^*) \prod_i A(u_i^*, u_{i+1}^*) = \text{sgn}(u') \prod_i A(u_i', u_{i+1}')$ , and similarly for  $v^*$  and  $v'$ . Finally, we sort the pairs so that the minimum elements increase, obtaining

$$u = 1, 10; 2, 9; 3, 8; 4, 11; 5, 12; 6, 7; \quad v = 1, 12; 2, 4; 3, 8; 5, 7; 6, 10; 9, 11.$$

This sorting does not introduce any further sign changes, so we have successfully transformed the term indexed by  $w$  in  $\det(A)$  to the term indexed by  $(u, v)$  in  $\text{Pf}(A)^2$ .

## 12.13 Domino Tilings of Rectangles

This section presents P. W. Kasteleyn's proof of a formula for the number of ways to tile a rectangle with dominos. Let  $\text{Dom}(m, n)$  be the set of domino tilings of a rectangle of width  $m$  and height  $n$ . This set is empty if  $m$  and  $n$  are both odd, so we will assume throughout that  $m$  is even. Given a tiling  $T \in \text{Dom}(m, n)$ , let  $N_h(T)$  and  $N_v(T)$  be the number of horizontal and vertical dominos (respectively) appearing in  $T$ . Define the *weight* of the tiling  $T$  to be  $\text{wt}(T) = x^{N_h(T)}y^{N_v(T)}$ .

**12.85. Theorem: Domino Tiling Formula.** For all even  $m \geq 1$  and all  $n \geq 1$ ,

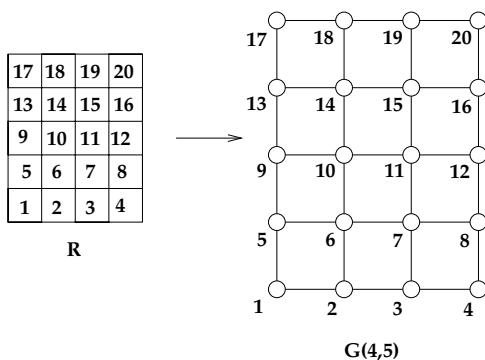
$$\sum_{T \in \text{Dom}(m, n)} \text{wt}(T) = 2^{mn/2} \prod_{j=1}^{m/2} \prod_{k=1}^n \sqrt{x^2 \cos^2 \left( \frac{j\pi}{m+1} \right) + y^2 \cos^2 \left( \frac{k\pi}{n+1} \right)}. \quad (12.12)$$

By setting  $x = y = 1$ , we obtain the expression for  $|\text{Dom}(m, n)|$  stated in the Introduction.

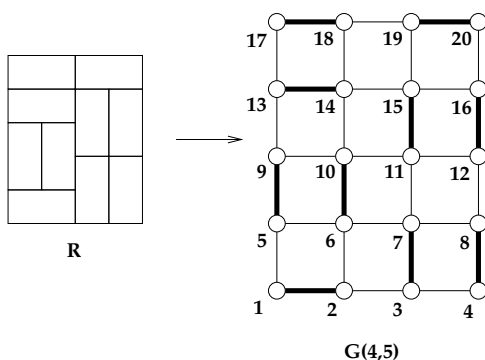
**Step 1: Conversion to a Perfect Matching Problem.** Introduce a simple graph  $G(m, n)$  with vertex set  $V = \{1, 2, \dots, mn\}$  and edge set  $E = E_x \cup E_y$ , where

$$E_x = \{\{k, k+1\} : k \not\equiv 0 \pmod{m}\}, \quad E_y = \{\{k, k+m\} : 1 \leq k \leq m(n-1)\}.$$

This graph models an  $m \times n$  rectangle  $R$ , as follows. The unit square in the  $i$ th row from the bottom and the  $j$ th column from the left in  $R$  corresponds to the vertex  $(i-1)m + j$ , for  $1 \leq i \leq n$  and  $1 \leq j \leq m$ . There is an edge in  $E_x$  for each pair of two horizontally adjacent

**FIGURE 12.15**

Graph used to model domino tilings.

**FIGURE 12.16**

A domino tiling and a perfect matching.

squares in  $R$ , and there is an edge in  $E_y$  for each pair of two vertically adjacent squares in  $R$ . There is a bijection between the set  $\text{Dom}(m, n)$  of domino tilings of  $R$  and the set  $\text{PM}(G(m, n))$  of perfect matchings of  $G(m, n)$ . Given a domino tiling, one need only replace each domino covering two adjacent squares by the edge associated to these two squares. This does give a perfect matching, since each square is covered by exactly one domino. If a tiling  $T$  corresponds to a matching  $M$  under this bijection, we have  $N_h(T) = |M \cap E_x|$  and  $N_v(T) = |M \cap E_y|$ . So, defining  $\text{wt}(M) = x^{|M \cap E_x|} y^{|M \cap E_y|}$ , we have

$$\sum_{T \in \text{Dom}(m, n)} \text{wt}(T) = \sum_{M \in \text{PM}(G(m, n))} \text{wt}(M).$$

**12.86. Example.** Figure 12.15 shows the rectangle  $R$  and associated graph  $G(m, n)$  when  $m = 4$  and  $n = 5$ . Figure 12.16 shows a domino tiling of  $R$  and the associated perfect matching. The tiling and matching shown both have weight  $x^4 y^6$ .

**Step 2: Enumeration via Pfaffians.** Let  $X_1$  be the skew-symmetric matrix defined in 12.81, taking  $G$  there to be  $G(m, n)$ . We know that

$$\sum_{M \in \text{PM}(G(m, n))} \text{sgn}(M) \prod_{\{i < j\} \in M} x_{i, j} = \text{Pf}(X_1). \quad (12.13)$$

$$\begin{bmatrix} 0 & x & 0 & 0 & -y & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -x & 0 & x & 0 & 0 & y & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -x & 0 & x & 0 & 0 & -y & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -x & 0 & 0 & 0 & 0 & y & 0 & 0 & 0 & 0 \\ y & 0 & 0 & 0 & 0 & x & 0 & 0 & -y & 0 & 0 & 0 \\ 0 & -y & 0 & 0 & -x & 0 & x & 0 & 0 & y & 0 & 0 \\ 0 & 0 & y & 0 & 0 & -x & 0 & x & 0 & 0 & -y & 0 \\ 0 & 0 & 0 & -y & 0 & 0 & -x & 0 & 0 & 0 & 0 & y \\ 0 & 0 & 0 & 0 & y & 0 & 0 & 0 & 0 & x & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -y & 0 & 0 & -x & 0 & x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & y & 0 & 0 & -x & 0 & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -y & 0 & 0 & -x & 0 \end{bmatrix}$$

**FIGURE 12.17**

Matrix used to enumerate domino tilings ( $m = 4, n = 3$ ).

We introduce the terms *horizontal edge*, *odd vertical edge*, and *even vertical edge* to refer (respectively) to edges in  $E_x$ , edges  $\{k, k + m\}$  in  $E_y$  with  $k$  odd, and edges  $\{k, k + m\}$  in  $E_y$  with  $k$  even. Consider the evaluation homomorphism (see 7.102) that sends  $x_{i,j}$  to  $x$  if  $\{i, j\}$  is a horizontal edge, sends  $x_{i,j}$  to  $y$  if  $\{i, j\}$  is an even vertical edge, and sends  $x_{i,j}$  to  $-y$  if  $\{i, j\}$  is an odd vertical edge. Let  $X$  be the matrix obtained by applying this homomorphism to each entry of the matrix  $X_1$ . Explicitly,  $X$  is the  $mn \times mn$  matrix with entries

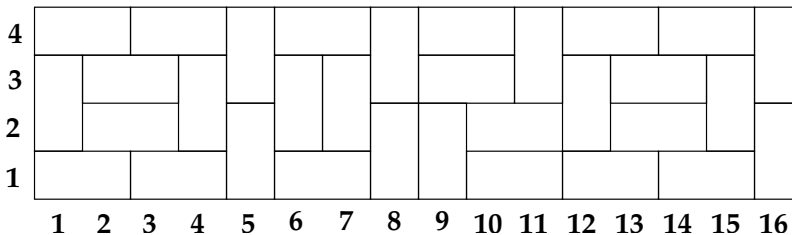
$$X(i, j) = \begin{cases} x & \text{if } j = i + 1 \text{ and } i \not\equiv 0 \pmod{m} \\ y & \text{if } j = i + m \text{ and } i \equiv 0 \pmod{2} \\ -y & \text{if } j = i + m \text{ and } i \equiv 1 \pmod{2} \\ -x & \text{if } i = j + 1 \text{ and } j \not\equiv 0 \pmod{m} \\ -y & \text{if } i = j + m \text{ and } j \equiv 0 \pmod{2} \\ y & \text{if } i = j + m \text{ and } j \equiv 1 \pmod{2} \\ 0 & \text{otherwise.} \end{cases} \quad (12.14)$$

For example, the matrix  $X$  when  $m = 4$  and  $n = 3$  appears in Figure 12.17. Let  $\text{sgn}^*(M) = \text{sgn}(M)(-1)^t$ , where  $t$  is the number of odd vertical edges in  $M$ . Applying the evaluation homomorphism to each side of (12.13) gives

$$\sum_{M \in \text{PM}(G(m,n))} \text{sgn}^*(M) \text{wt}(M) = \text{Pf}(X).$$

**Step 3: Sign Analysis.** The crucial fact to be verified is that  $\text{sgn}^*(M) = +1$  for every  $M$ . Before proving this fact, we consider an example.

**12.87. Example.** Consider the following domino tiling of a  $16 \times 4$  rectangle:



This tiling corresponds to a perfect matching  $M$  of  $G(16, 4)$ , which is encoded (via 12.80) by a word  $w \in \text{SPf}_{64}$ . By definition,  $\text{sgn}(M) = (-1)^{\text{inv}(w)}$ . In our example, the word of  $M$  is

$$w = 1, 2; 3, 4; 5, 21; 6, 7; 8, 24; 9, 25; 10, 11; 12, 13; 14, 15; 16, 32; \\ 17, 33; 18, 19; 20, 36; 22, 38; 23, 39; 26, 27; 28, 44; 29, 30; 31, 47; \dots; 60, 61; 62, 63.$$

Note that  $w$  consists of pairs of letters indicating the two squares occupied by each domino in the tiling. We imagine placing dominos on the board one at a time, in the order specified by  $w$ , and updating  $\text{sgn}(M)$  and  $\text{sgn}^*(M)$  as we go along. When computing  $\text{inv}(w)$ , the second symbol in each pair sometimes causes inversions with symbols following it in  $w$ . Pairs corresponding to horizontal dominos never cause any inversions. Consider the inversions caused by a vertical domino (i.e., a vertical edge in  $M$ ). The first vertical edge appearing in  $w$  is  $\{5, 21\}$ . The 21 is greater than the fifteen symbols  $6, 7, \dots, 20$  corresponding to squares to the right of column 5 in row 1 and squares to the left of column 5 in row 2, which have not been covered by a domino yet. So this edge increases  $\text{inv}(w)$  by  $15 = m - 1$ , which causes a sign change in  $\text{sgn}(M)$ . However, since this edge is an odd vertical edge, that sign change is counteracted in  $\text{sgn}^*(M)$ .

The next vertical edge in  $w$  is  $\{8, 24\}$ . The symbol 24 causes  $14 = m - 2$  new inversions, corresponding to squares to the right of column 8 in row 1 and squares to the left of column 8 in row 2, excluding column 5. These inversions do not change  $\text{sgn}(M)$ , and  $\text{sgn}^*(M)$  is also unchanged since  $\{8, 24\}$  is an even vertical edge.

Continuing similarly, we eventually come to the odd vertical edge  $\{23, 39\}$  in  $w$ . Recalling the order of domino placement, we see that the 39 causes inversions with the following nine symbols to its right in  $w$ : 37, 35, 34, 31, 30, 29, 28, 27, 26. Since nine is odd, we get a sign change in  $\text{sgn}(M)$ , but this is counteracted in  $\text{sgn}^*(M)$  since we have just added an odd vertical edge. After accounting for all the dominos, we find that indeed  $\text{sgn}^*(M) = +1$ , since the insertion of each vertical domino never leads to a net sign change (see 12.170).

Now we are ready to prove that  $\text{sgn}^*(M) = +1$  for a general  $M \in G(m, n)$ . Let  $w \in \text{SPf}_{mn}$  be the word encoding  $M$ . As in the example, we calculate  $\text{sgn}^*(M) = (-1)^{\text{inv}(w)}(-1)^t$  incrementally by scanning the edges in  $w$  from left to right. Initially, before scanning any edges, this quantity is  $+1$ . Suppose the next edge in the scan is the horizontal edge  $\{k, k+1\}$ . By definition of  $w$  (see 12.80),  $k$  is the smallest symbol that has not appeared previously in  $w$ . So  $k$  and  $k+1$  cannot cause any new inversions with symbols following them. Similarly,  $t$  (the number of odd vertical edges) does not increase when we scan this edge. So  $\text{sgn}^*(M)$  is still  $+1$  after scanning this edge.

Before continuing, we need the following observation: for every row  $i \geq 1$ , the number of vertical dominos that start in row  $i$  and end in row  $i+1$  is even (possibly zero). This is proved by induction on  $i$ . To prove the case  $i = 1$ , suppose there are  $a$  horizontal dominos in row 1. Then there must be  $m - 2a$  vertical dominos starting in row 1. This number is even, since  $m$  is even. Now assume the result holds in row  $i-1$ . In row  $i$ , suppose there are  $a$  horizontal dominos,  $b$  vertical dominos coming up from row  $i-1$ , and  $c$  vertical dominos leading up into row  $i+1$ . Then  $c = m - 2a - b$ . Since  $m$  is even and (by hypothesis)  $b$  is even,  $c$  must also be even.

Now suppose the next edge in the scan is a vertical edge  $\{k, k+m\}$  in column  $j$  that covers rows  $i$  and  $i+1$  (so  $k = (i-1)m+j$ ). As before, the symbol  $k$  causes no new inversions. Let us count the inversions in  $w$  between  $k+m$  and symbols to its right. There are  $m-1$  symbols that might cause inversions with  $k+m$ , namely  $k+1, k+2, \dots, k+(m-1)$ , but some of these symbols may have already appeared in  $w$ . Specifically, if there are  $a$  vertical dominos covering rows  $i$  and  $i+1$  to the left of column  $j$ , and  $b$  vertical dominos covering rows  $i-1$  and  $i$  to the right of column  $j$ , then  $a+b$  of the symbols just mentioned will have

already appeared in  $w$ . So, the inclusion of the new edge increases  $\text{inv}(w)$  by  $(m-1)-(a+b)$ . Now, let there be  $b'$  vertical dominos covering rows  $i-1$  and  $i$  to the left of column  $j$ , and  $c$  horizontal dominos in row  $i$  to the left of column  $j$ . Since  $m-1 \equiv 1 \pmod{2}$ ,  $-a \equiv a \pmod{2}$ ,  $-b \equiv b' \pmod{2}$  (by the observation in the last paragraph), and  $2c \equiv 0 \pmod{2}$ , we see that

$$(m-1) - a - b \equiv 1 + a + b' + 2c \pmod{2}.$$

But  $1 + a + b' + 2c = j$  since  $a + b' + 2c$  counts all the columns left of column  $j$  in row  $i$ . We conclude, finally, that the increase in  $\text{inv}(w)$  caused by the insertion of the edge  $\{k, k+m\}$  has the same parity as the column index  $j$ . Since  $j$  and  $k$  have the same parity, the number of new inversions is odd iff the new vertical edge is an odd vertical edge. So there is no net change in  $\text{sgn}(M^*) = (-1)^{\text{inv}(w)}(-1)^t$  when we add this edge. This completes the proof that  $\text{sgn}(M^*) = +1$ .

**Step 4: Evaluation of the Pfaffian.** Combining steps 1 through 3 and 12.83, we have

$$\sum_{T \in \text{Dom}(m,n)} \text{wt}(T) = \sum_{M \in \text{PM}(G(m,n))} \text{wt}(M) = \text{Pf}(X) = \sqrt{\det(X)},$$

where  $X$  is the  $mn \times mn$  matrix defined by (12.14). So we are reduced to evaluating the determinant of  $X$ . The idea is to replace  $X$  by a similar matrix  $U^{-1}XU$  whose determinant is easier to evaluate. For this purpose, it is convenient to introduce tensor products of matrices.

**12.88. Definition: Tensor Product of Matrices.** If  $A$  is any  $n \times n$  matrix and  $B$  is any  $m \times m$  matrix, let  $A \otimes B$  be the  $mn \times mn$  matrix given in block form by

$$A \otimes B = \begin{bmatrix} a_{1,1}B & a_{1,2}B & \cdots & a_{1,n}B \\ a_{2,1}B & a_{2,2}B & \cdots & a_{2,n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{n,1}B & a_{n,2}B & \cdots & a_{n,n}B \end{bmatrix}.$$

Formally,  $(A \otimes B)(m(i_1 - 1) + i_2, m(j_1 - 1) + j_2) = A(i_1, j_1)B(i_2, j_2)$  for all  $1 \leq i_1, j_1 \leq n$  and all  $1 \leq i_2, j_2 \leq m$ .

The following properties of tensor products may be routinely verified:

- (a)  $(A_1 + A_2) \otimes B = (A_1 \otimes B) + (A_2 \otimes B)$  and  $A \otimes (B_1 + B_2) = (A \otimes B_1) + (A \otimes B_2)$ .
- (b) For any scalar  $c$ ,  $(cA) \otimes B = c(A \otimes B) = A \otimes (cB)$ .
- (c)  $(A_1 \otimes B_1)(A_2 \otimes B_2) = (A_1 A_2) \otimes (B_1 B_2)$ .
- (d) If  $A$  and  $B$  are invertible, then  $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ .

For every  $k \geq 1$ , let  $I_k$  denote the  $k \times k$  identity matrix, let  $F_k$  denote the  $k \times k$  diagonal matrix with diagonal entries  $-1, 1, -1, 1, \dots, (-1)^k$ , let  $I'_k$  denote the  $k \times k$  matrix with 1's on the antidiagonal, and let  $Q_k$  denote the  $k \times k$  matrix with ones on the diagonal above the main diagonal,  $-1$ 's on the diagonal below the main diagonal, and zeroes elsewhere. For example,

$$I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad F_5 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix},$$

$$I'_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad Q_5 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix}.$$

$$2i \begin{bmatrix} xr_1 & 0 & 0 & -ys_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & xr_2 & -ys_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -ys_1 & xr_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -ys_1 & 0 & 0 & xr_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & xr_1 & 0 & 0 & -ys_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & xr_2 & -ys_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -ys_2 & xr_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -ys_2 & 0 & 0 & xr_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & xr_1 & 0 & 0 & -ys_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & xr_2 & -ys_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -ys_3 & xr_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -ys_3 & 0 & 0 & xr_4 \end{bmatrix}$$

**FIGURE 12.18**

Transformed matrix  $U^{-1}XU$  for  $m = 4$ ,  $n = 3$ . (Here  $r_a = 2i \cos(\pi a/5)$  and  $s_b = 2i \cos(\pi b/4)$ .)

The definition of  $X$  in (12.14) can now be written

$$X = x(I_n \otimes Q_m) + y(Q_n \otimes F_m).$$

(Compare to Figure 12.17.) The following lemma can be established by routine calculations, which we leave to the reader.

**12.89. Lemma: Eigenvectors of  $Q_k$ .** For  $0 \leq a \leq k+1$  and  $1 \leq b \leq k$ , define complex numbers

$$U_k(a, b) = i^a \sin\left(\frac{\pi ab}{k+1}\right), \quad \lambda_k(b) = 2i \cos\left(\frac{b\pi}{k+1}\right).$$

For  $1 \leq a, b \leq k$ , we have

$$U_k(a+1, b) - U_k(a-1, b) = \lambda_k(b)U_k(a, b).$$

Therefore, the column vector  $(U_k(1, b), U_k(2, b), \dots, U_k(a, b))^t$  is an eigenvector of  $Q_k$  associated to the eigenvalue  $\lambda_k(b)$ . Letting  $U_k = (U_k(a, b))_{1 \leq a, b \leq k}$  and  $D_k$  be the  $k \times k$  diagonal matrix with diagonal entries  $\lambda_k(b)$ , we have  $Q_k U_k = U_k D_k$ . Furthermore,  $(-1)^a U_k(a, b) = -U_k(a, k+1-b)$  for  $1 \leq a, b \leq k$ , and therefore  $F_k U_k = -U_k I'_k$ .

The columns of  $U_k$  are linearly independent, because they are eigenvectors of  $Q_k$  associated to *distinct* eigenvalues. Therefore,  $U_k$  is invertible, so the lemma gives  $U_k^{-1} Q_k U_k = D_k$  and  $U_k^{-1} F_k U_k = -I'_k$ . Let  $U = U_n \otimes U_m$ , so  $U^{-1} = U_n^{-1} \otimes U_m^{-1}$ . Using properties of tensor products, we calculate

$$\begin{aligned} U^{-1}XU &= x(U_n^{-1} \otimes U_m^{-1})(I_n \otimes Q_m)(U_n \otimes U_m) + y(U_n^{-1} \otimes U_m^{-1})(Q_n \otimes F_m)(U_n \otimes U_m) \\ &= x(U_n^{-1} I_n U_n) \otimes (U_m^{-1} Q_m U_m) + y(U_n^{-1} Q_n U_n) \otimes (U_m^{-1} F_m U_m) \\ &= x(I_n \otimes D_m) - y(D_n \otimes I'_m). \end{aligned}$$

For example, if  $X$  is the matrix shown in Figure 12.17, then  $U^{-1}XU$  is the matrix shown in Figure 12.18. In general,  $U^{-1}XU$  is a block-diagonal matrix consisting of  $n$   $m \times m$  blocks. The  $b$ th block has entries  $-y\lambda_n(b)$  on the anti-diagonal and entries  $x\lambda_m(a)$  (for  $1 \leq a \leq m$ ) on the diagonal. Now, since  $m$  is even, we can reorder the rows and columns of each block

into this order:  $1, m, 2, m-1, 3, m-2, \dots, m/2, m/2+1$ . This reordering can be accomplished by performing an even number of row and column switches on  $U^{-1}XU$ , so the determinant does not change. The new matrix is also block-diagonal, consisting of  $(mn/2)$   $2 \times 2$  blocks that look like

$$\begin{bmatrix} x\lambda_m(a) & -y\lambda_n(b) \\ -y\lambda_n(b) & x\lambda_m(m+1-a) \end{bmatrix} \quad (1 \leq a \leq m/2, 1 \leq b \leq n).$$

Now,  $\lambda_m(m+1-a) = 2i \cos(\pi(m+1-a)/(m+1)) = -2i \cos(\pi a/(m+1)) = -\lambda_m(a)$ . It follows that the determinant of the  $2 \times 2$  block just mentioned is

$$-x^2\lambda_m(a)^2 - y^2\lambda_n(b)^2 = 4 \left[ x^2 \cos^2 \left( \frac{\pi a}{m+1} \right) + y^2 \cos^2 \left( \frac{\pi b}{n+1} \right) \right].$$

Finally,  $\det(X) = \det(U^{-1}XU)$  is the product of these determinants as  $a$  ranges from 1 to  $m/2$  and  $b$  ranges from 1 to  $n$ . Taking the square root of  $\det(X)$  and factoring out powers of 2 produces formula (12.12).

## Summary

- *Rational-Slope Dyck Paths.* If  $\gcd(r, s) = 1$ , then the number of lattice paths from  $(0, 0)$  to  $(r, s)$  that never go below the line  $sx = ry$  is  $\frac{1}{r+s} \binom{r+s}{r, s}$ . For any lattice path ending at  $(r, s)$ , the  $r + s$  cyclic shifts of this path are all distinct, and exactly one of them is an  $r/s$ -Dyck path.
- *Chung-Feller Theorem.* A path from  $(0, 0)$  to  $(n, n)$  has  $k$  flaws iff the path has  $k$  north steps starting below  $y = x$ . For  $0 \leq k \leq n$ , there are  $C_n = \frac{1}{n+1} \binom{2n}{n, n}$  paths ending at  $(n, n)$  with  $k$  flaws. Thus the number of flaws in a random path is uniformly distributed on  $\{0, 1, 2, \dots, n\}$ .
- *Rook-Equivalence of Ferrers Boards.* For each integer partition  $\mu$ ,  $r_k(\mu)$  is the number of ways to place  $k$  non-attacking rooks on  $F_\mu = \text{dg}(\mu)$ , and  $R_\mu(x) = \sum_{k \geq 0} r_k(\mu) x^k$ . For all partitions  $\mu = (\mu_1 \geq \mu_2 \geq \dots \geq \mu_n \geq 0)$  and  $\nu = (\nu_1 \geq \nu_2 \geq \dots \geq \nu_n \geq 0)$  with  $|\mu| = n = |\nu|$ , we have  $R_\mu(x) = R_\nu(x)$  iff the multisets  $[\mu_i + i : 1 \leq i \leq n]$  and  $[\nu_i + i : 1 \leq i \leq n]$  are equal.
- *Parking Functions.* A function  $f : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  is a *parking function* iff  $|\{x : f(x) \leq i\}| \geq i$  for all  $i \leq n$ . There are  $(n+1)^{n-1}$  parking functions of order  $n$ . A bijection from parking functions to labeled Dyck paths is given by putting the labels  $\{x : f(x) = i\}$  in increasing order in column  $i$  for each  $i$ . A bijection from labeled Dyck paths to trees is given by letting the children of  $a_i$  be the labels in column  $i+1$ , for all  $i \geq 0$  (where  $a_0 = 0$  and  $a_1, \dots, a_n$  are the labels from bottom to top).
- *Facts about Cyclic Groups.* If  $G$  is a cyclic group of size  $n < \infty$ , then  $G$  has a unique cyclic subgroup of size  $d$  for each divisor  $d$  of  $n$ , and these are all the subgroups of  $G$ . Any cyclic group of size  $d$  has  $\phi(d)$  generators, and hence  $n = \sum_{d|n} \phi(d)$ . If  $G$  is a group of size  $n$  with at most one subgroup of size  $d$  for each divisor  $d$  of  $n$ , then  $G$  must be cyclic. Hence, any finite subgroup of the multiplicative group of a field is cyclic.
- *Counting Irreducible Polynomials.* The size of a finite field must be a prime power. For each prime power  $q$ , there exists a field  $F$  with  $q$  elements, which is unique up



to isomorphism. For such a field  $F$ , let  $I(n, q)$  be the number of monic irreducible polynomials of degree  $n$  in  $F[x]$ . Classifying elements in the field of size  $q^n$  by their minimal polynomials in  $F[x]$  gives  $q^n = \sum_{d|n} dI(d, q)$ . Hence, by Möbius inversion,  $I(n, q) = \frac{1}{n} \sum_{d|n} q^d \mu(n/d)$  where  $\mu$  is the Möbius function defined in 4.28.

- *Subspaces of Vector Spaces over Finite Fields.* A  $d$ -dimensional vector space over a  $q$ -element field has size  $q^d$ . The number of  $k$ -dimensional subspaces of an  $n$ -dimensional vector space over a  $q$ -element field is  $\begin{bmatrix} n \\ k \end{bmatrix}_q$ . Each such subspace has a unique basis in reduced row-echelon form (RREF). The number of  $k \times n$  RREF matrices with entries in a  $q$ -element field is thus  $\begin{bmatrix} n \\ k \end{bmatrix}_q$ .
- *Combinatorial Meaning of Tangent and Secant Power Series.*  $\tan x = \sum_{n \geq 0} (a_n/n!)x^n$ , where  $a_n$  counts permutations  $w$  satisfying  $w_1 < w_2 > w_3 < w_4 > \cdots > w_n$ ; and  $\sec x = \sum_{n \geq 0} (b_n/n!)x^n$ , where  $b_n$  counts permutations  $w$  satisfying  $w_1 < w_2 > w_3 < \cdots < w_n$ .
- *Tournaments.* A tournament is a digraph with exactly one directed edge between each pair of distinct vertices. A tournament  $t$  is transitive iff  $(u, v) \in t$  and  $(v, w) \in t$  always imply  $(u, w) \in t$  iff  $t$  contains no directed 3-cycle iff the outdegrees of the vertices of  $t$  are pairwise distinct. A sign-reversing involution exists that cancels all non-transitive tournaments, leading to this formula for the Vandermonde determinant:

$$\det \|x_j^{n-i}\|_{1 \leq i, j \leq n} = \sum_{w \in S_n} \operatorname{sgn}(w) \prod_{k=1}^n x_{w(k)}^{n-k} = \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

- *Hook-Length Formula.* For a partition  $\lambda$  with  $n$  boxes, the number of standard tableaux of shape  $\lambda$  is  $n! / \prod_{c \in \operatorname{dg}(\lambda)} h(c)$ , where  $h(c)$  is the hook-length of cell  $c$ . This can be proved probabilistically by defining a random algorithm that generates each  $S \in \operatorname{SYT}(\lambda)$  with probability  $\prod_{c \in \operatorname{dg}(\lambda)} h(c)/n!$ . To build  $S$ , start at a random cell in  $\operatorname{dg}(\lambda)$ , then repeatedly jump to a random cell in the hook of the current cell until reaching a corner. Place  $n$  in this corner and proceed recursively to fill the other cells in  $\operatorname{dg}(\lambda)$ .
- *Knuth Equivalence and Monotone Subsequences of Words.* Two words  $v$  and  $w$  are Knuth equivalent iff  $v$  can be changed into  $w$  by a sequence of moves of the form  $\cdots yxz \cdots \leftrightarrow \cdots yzx \cdots$  (where  $x < y \leq z$ ) or  $\cdots xzy \cdots \leftrightarrow \cdots zxy \cdots$  (where  $x \leq y < z$ ). These moves simulate tableau insertion (when applied to reading words), so every  $w$  is Knuth equivalent to the reading word of its insertion tableau  $P(w)$ . Words  $v$  and  $w$  are Knuth equivalent iff  $P(v) = P(w)$ . If  $P(w)$  has shape  $\lambda$ , then  $\lambda_1 + \cdots + \lambda_k$  is the maximum total length of a set of  $k$  disjoint weakly increasing subsequences of  $w$ , and  $\lambda'_1 + \cdots + \lambda'_k$  is the maximum total length of a set of  $k$  disjoint strictly decreasing subsequences of  $w$ .
- *Pfaffians.* Let  $N$  be even. Given an  $N \times N$  matrix  $A$  that is skew-symmetric ( $A^t = -A$ ), the Pfaffian of  $A$  is

$$\operatorname{Pf}(A) = \sum_{w \in \operatorname{SPf}_N} \operatorname{sgn}(w) \prod_{i \text{ odd}} A(w_i, w_{i+1}),$$

where  $w \in \operatorname{SPf}_N$  iff  $w \in S_N$ ,  $w_i < w_{i+1}$ , and  $w_i < w_{i+2}$  for all odd  $i$ . We have  $\det(A) = \operatorname{Pf}(A)^2$ . Each term of  $\operatorname{Pf}(A)$  counts a signed, weighted perfect matching of a graph with vertex set  $\{1, 2, \dots, N\}$ , where an edge from  $i$  to  $j$  (for  $i < j$ ) is weighted by  $A(i, j)$ . There is a recursion  $\operatorname{Pf}(A) = \sum_{j=2}^N (-1)^j A(1, j) \operatorname{Pf}(A[[1, j]])$ , where  $A[[1, j]]$  is the matrix obtained by deleting rows 1 and  $j$  and columns 1 and  $j$  from  $A$ .

- *Domino Tilings.* For all  $m, n \in \mathbb{N}^+$  with  $m$  even, the coefficient of  $x^a y^b$  in

$$2^{mn/2} \prod_{j=1}^{m/2} \prod_{k=1}^n \sqrt{x^2 \cos^2 \left( \frac{j\pi}{m+1} \right) + y^2 \cos^2 \left( \frac{k\pi}{n+1} \right)}$$

is the number of ways to tile an  $m \times n$  board with  $a$  horizontal dominos and  $b$  vertical dominos. The steps in the proof are: (a) model domino tilings by perfect matchings of a grid-shaped graph; (b) use a Pfaffian to enumerate these signed perfect matchings; (c) adjust signs in the matrix so every perfect matching has sign  $+1$ ; (d) rewrite the Pfaffian as the square root of the determinant of the matrix; (e) evaluate the determinant by performing a similarity transformation that nearly diagonalizes the matrix, creating  $2 \times 2$  blocks running down the diagonal. Each  $2 \times 2$  block contributes one of the factors in the product formula above.

## Exercises

**12.90.** Let  $\sim$  be the cyclic shift relation from §12.1. Find all the equivalence classes of  $\sim$  for: (a) the set of lattice paths ending at  $(3, 4)$ ; (b) the set of lattice paths ending at  $(3, 3)$ .

**12.91.** For  $v, w \in \mathcal{R}(N^s E^r)$ , write  $v \sim w$  iff  $w$  can be obtained from  $v$  by a cyclic shift. Which of the following statements is always true for all  $r, s \geq 1$ ? (a) Every equivalence class of  $\sim$  has size  $r + s$ . (b) Every equivalence class of  $\sim$  contains at least one  $r/s$ -Dyck path. (c) Every equivalence class of  $\sim$  contains at most one  $r/s$ -Dyck path.

**12.92.** Let  $k \geq 0$  and  $m \geq 1$  be integers. Show that the number of lattice paths from  $(0, 0)$  to  $(k + mh, h)$  that never go below the line  $x = k + my$  is

$$\binom{k + (m+1)h}{k + mh, h} - m \binom{k + (m+1)h}{k + mh + 1, h - 1}.$$

Give a bijective proof analogous to the proof of 1.56 in §1.10.

**12.93.** Verify the Chung-Feller theorem directly for  $n = 3$  by drawing all lattice paths from  $(0, 0)$  to  $(3, 3)$  with: (a) 0 flaws; (b) 1 flaw; (c) 2 flaws; (d) 3 flaws.

**12.94.** Let  $\pi$  be the Dyck path NNENEENNENNNENNEEENENEE. Use the bijections from 12.4 to compute the associated lattice path with: (a) 5 flaws; (b) 8 flaws; (c) 10 flaws.

**12.95.** For each flawed path  $\pi$ , find the Dyck path associated to  $\pi$  via the bijections in 12.4: (a) NENNEEEENENNNNEEENEENNNNE; (b) NEEENNENEEENNNNNEENE.

**12.96.** Let  $\pi$  be a random lattice path from  $(0, 0)$  to  $(n, n)$ , and for  $1 \leq j \leq n$ , let

$$X_j(\pi) = \chi(\pi \text{ has a flaw in row } j).$$

Prove bijectively that  $P(X_j = 0) = 1/2 = P(X_j = 1)$ .

**12.97.** Let  $X_1, X_2, \dots, X_n$  be *independent* random variables such that  $P(X_i = 1) = 1/2 = P(X_i = 0)$  for all  $i$ . (This means that, for all  $v_1, \dots, v_n \in \{0, 1\}$ , the events  $X_1 = v_1, X_2 = v_2, \dots, X_n = v_n$  are independent in the sense of 1.84.) Compute  $P(X_1 + X_2 + \dots + X_n = k)$  for  $0 \leq k \leq n$ . Contrast your answer with the Chung-Feller theorem.

**12.98.** Find a formula for the number of lattice paths from  $(0, 0)$  to  $(n, n)$  with  $k$  flaws and  $j$  east steps departing from the line  $y = x$ .

**12.99.** Compute the rook polynomial for each of the following partitions:

(a)  $(3, 2, 1)$ ; (b)  $(8, 8, 8, 8, 8, 8, 8, 8)$ ; (c)  $(n)$ ; (d)  $(n, n, 1^k)$ .

**12.100.** Draw the diagrams of all integer partitions of 8 and determine which pairs of partitions are rook-equivalent.

**12.101.** Prove: for any integer partition  $\mu$ ,  $R_\mu(x) = R_{\mu'}(x)$ .

**12.102.** (a) For any  $n \geq 1$ , prove that the partition  $\mu$  consisting of  $n$  copies of  $n$  is rook-equivalent to the partition  $\nu = (2n - 1, 2n - 3, \dots, 5, 3, 1)$ . (b) Define a bijection between the set of non-attacking placements of  $k$  rooks on  $\mu$  and the set of non-attacking placements of  $k$  rooks on  $\nu$ .

**12.103.** Let  $\mu$  be an integer partition such that  $\text{dg}(\mu) \subseteq \text{dg}(\Delta_N)$ , where  $\Delta_N = (N - 1, N - 2, \dots, 3, 2, 1, 0)$ . Suppose the sequence  $(N - 1 - \mu_1, N - 2 - \mu_2, \dots, 0 - \mu_N)$  has  $a_k$  copies of  $k$  for  $k \geq 0$ . (Note that this sequence gives the row lengths of the skew shape  $\Delta_N / \mu$ .) Prove that the number of partitions that are rook-equivalent to  $\mu$  is

$$\prod_{k \geq 1} \binom{a_{k-1} + a_k - 1}{a_{k-1} - 1, a_k}.$$

**12.104.** Show that for each integer partition  $\mu$ , there is a unique integer partition  $\nu$  with distinct parts that is rook-equivalent to  $\mu$ .

**12.105.** Suppose  $\mu$  is an integer partition with  $\text{dg}(\mu) \subseteq \text{dg}(\Delta_N)$ , where  $\Delta_N = (N - 1, N - 2, \dots, 2, 1, 0)$ . (a) Using a suitable involution, prove that

$$r_k(\mu) = \sum_{i=0}^k S(N - i, N - k) (-1)^i e_i(N - 1 - \mu_1, N - 2 - \mu_2, \dots, N - N - \mu_N),$$

where  $S(u, v)$  is a Stirling number of the second kind and  $e_i$  is an elementary symmetric polynomial. (b) Deduce from (a) a combinatorial proof of part (d) of 2.77. (c) Deduce from (a) that the multiset condition in 12.10 is sufficient for  $R_\mu(x) = R_\nu(x)$ . (d) Assume  $\mu$  and  $\nu$  are rook-equivalent partitions. Use (a) and the Garsia-Milne involution principle 4.126 to construct a bijection from the set of non-attacking placements of  $k$  rooks on  $F_\mu$  to the set of non-attacking placements of  $k$  rooks on  $F_\nu$ .

**12.106.** For each labeled Dyck path in Figure 12.9, compute the associated parking function and tree (see 12.21 and 12.22).

**12.107.** (a) Convert the parking function  $f$  in Figure 12.5 to a labeled Dyck path and a tree. (b) Convert the labeled Dyck path NNENNEENEENENNEE with labels 5, 8, 2, 4, 1, 6, 3, 7 (from bottom to top) to a parking function and a tree. (c) Convert the tree

$$T = (\{0, 1, \dots, 10\}, \{\{0, 9\}, \{5, 7\}, \{5, 8\}, \{9, 4\}, \{7, 6\}, \{6, 9\}, \{7, 10\}, \{10, 1\}, \{3, 9\}, \{2, 9\}\})$$

to a labeled Dyck path and a parking function.

**12.108.** Suppose we represent a function  $f : \{1, 2, \dots, b\} \rightarrow \{1, 2, \dots, a + 1\}$  as a labeled lattice path ending at  $(a, b)$ . Find conditions on the labeled path that are equivalent to  $f$  being (a) surjective; (b) injective.

**12.109.** (a) Given nonnegative integers  $c_1, \dots, c_{a+1}$  adding to  $b$ , how many labeled lattice paths from  $(0, 0)$  to  $(a, b)$  have  $c_i$  labels in column  $i$  for all  $i$ ? (b) Use the bijections in §12.5 to translate (a) into enumeration results for parking functions and trees.

**12.110.** (a) Let  $p_n$  be the number of parking functions of order  $n$ . Give a combinatorial proof of the recursion

$$p_n = \sum_{m=1}^n m \binom{n-1}{m-1} p_{m-1} p_{n-m}.$$

(b) Use (a) and 3.186 to define a bijection between parking functions and trees.

**12.111.** For a parking function  $f \in \mathcal{P}_n$ , let  $\text{wt}(f) = n(n+1)/2 - \sum_{i=1}^n f(i)$ . Let  $P_n(x) = \sum_{f \in \mathcal{P}_n} x^{\text{wt}(f)}$ . Prove the recursion

$$P_n(x) = \sum_{m=1}^n [m]_x \binom{n-1}{m-1} P_{m-1}(x) P_{n-m}(x).$$

**12.112.** Let  $S$  be a  $k$ -element subset of  $\{1, 2, \dots, n\}$ . Prove that there are  $kn^{n-k-1}$  parking functions  $f$  such that  $S = \{x : f(x) = 1\}$ .

**12.113.** For each  $n, k, m \in \mathbb{N}$ , let  $\mathcal{P}_{n,k,m}$  be the set of labeled lattice paths ending at  $(k+mn, n)$  that never go below the line  $x = k + my$ . Find a recursion satisfied by the quantities  $|\mathcal{P}_{n,k,m}|$ .

**12.114.** Find a bijection between the set of parking functions of order  $n$  and the quotient group  $\mathbb{Z}_{n+1}^n/H$ , where  $H$  is the subgroup generated by  $(1, 1, \dots, 1)$ .

**12.115.** How many generators does an infinite cyclic group have?

**12.116.** Prove or disprove: if every proper subgroup of a finite group  $G$  is cyclic, then  $G$  itself must be cyclic.

**12.117.** Suppose  $G$  is a group such that, for all  $d \geq 1$ ,  $G$  has at most  $d$  elements  $x$  such that  $x^d = 1$ . Prove that every finite subgroup of  $G$  is cyclic.

**12.118.** Describe all the finite subgroups of the field  $\mathbb{C}$ .

**12.119. Quaternions.** Let  $H$  be a four-dimensional real vector space with basis  $1, i, j, k$ . Define multiplication on  $H$  by letting  $1$  act as the identity, setting  $i^2 = j^2 = k^2 = -1$ ,  $ij = k = -ji$ ,  $jk = i = -kj$ ,  $ki = j = -ik$ , and extending by linearity. (a) Show that  $H$  with this multiplication is a division ring (i.e.,  $H$  satisfies all the axioms in the definition of a field except commutativity of multiplication). (b) Find a non-cyclic finite subgroup of  $H^*$  (cf. 12.29). (c) Show that the equation  $x^2 = -1$  has infinitely many solutions in  $H^*$ .

**12.120.** Prove that the product of all the nonzero elements in a finite field  $F$  is  $-1_F$ . Deduce Wilson's theorem: for  $p$  prime,  $(p-1)! \equiv -1 \pmod{p}$ .

**12.121.** Compute the number of monic irreducible polynomials of degree 12 over a 9-element field.

**12.122.** (a) Enumerate all the irreducible polynomials in  $\mathbb{Z}_2[x]$  of degree at most 5. (b) Use the formula in 12.30 to compute  $I(n, 2)$  for  $1 \leq n \leq 8$  (compare with the results in (a) for  $n \leq 5$ ).

**12.123. Construction of Finite Fields.** Let  $F$  be a field with  $q$  elements, let  $h \in F[x]$  be a fixed monic irreducible polynomial of degree  $n$ , and let

$$K = \{f \in F[x] : f = 0 \text{ or } \deg(f) < n\}.$$

For  $f, g \in K$ , define  $f + g$  to be the usual sum of polynomials in  $F[x]$ , and define  $f \times g$  to be the remainder when  $fg$  is divided by  $h$ . Show that  $K$ , with these operations, is a field of size  $q^n$ . The field  $K$  is denoted  $F[x]/(h)$ .

**12.124.** Let  $K = \mathbb{Z}_2[x]/(x^3 + x + 1)$  (see 12.123). Construct addition and multiplication tables for  $K$ . Explicitly confirm that  $K^*$  is generated by  $x$  by computing  $x^i$  for  $1 \leq i \leq 7$ .

**12.125.** Let  $h = x^4 + x + 1 \in \mathbb{Z}_2[x]$ , and let  $K = \mathbb{Z}_2[x]/(h)$ , which is a 16-element field (see 12.123). (a) Explain why every element  $y \in K$  satisfies  $y^{16} = y$ . (b) List all the elements of  $K$  and their minimal polynomials over  $\mathbb{Z}_2$ . (c) Factor the polynomial  $x^{16} - x \in \mathbb{Z}_2[x]$  into a product of irreducible polynomials. (d) Explain the relation between part (b), part (c), and the formulas in 12.30. (e) Find all generators of the cyclic group  $K^*$ .

**12.126.** (a) Use 12.30 to show that  $I(n, q) > 0$  for all prime powers  $q$  and all  $n \geq 1$ . (b) Prove that for every prime power  $p^n$ , there exists a field of size  $p^n$ .

**12.127.** Let  $F$  be a finite field of size  $q$ . A polynomial  $h \in F[x]$  is called *primitive* iff  $h$  is a monic irreducible polynomial such that  $x$  is a generator of the multiplicative group of the field  $K = F[x]/(h)$  (see 12.123). (a) Count the primitive polynomials of degree  $n$  in  $F[x]$ . (b) Give an example of an irreducible polynomial in  $\mathbb{Z}_2[x]$  that is not primitive.

**12.128.** Let  $K$  be a  $q$ -element field. How many  $n \times n$  matrices with entries in  $K$  are: (a) upper-triangular; (b) strictly upper-triangular (zeroes on the main diagonal); (c) unitriangular (ones on the main diagonal); (d) upper-triangular and invertible?

**12.129.** How many  $2 \times 2$  matrices with entries in a  $q$ -element field have determinant 1?

**12.130.** Count the number of invertible  $n \times n$  matrices with entries in a  $q$ -element field  $F$ . How is the answer related to  $[n]_q!$ ?

**12.131.** How many 3-dimensional subspaces does the vector space  $\mathbb{Z}_7^5$  have?

**12.132.** For each integer partition  $\mu$  that fits in a box with 2 rows and 3 columns, draw a picture of the RREF matrix associated to  $\mu$  in the proof of 12.37.

**12.133.** Find the RREF basis for the subspace of  $\mathbb{Z}_5^5$  spanned by  $v_1 = (1, 4, 2, 3, 4)$ ,  $v_2 = (2, 3, 1, 0, 0)$ , and  $v_3 = (0, 0, 3, 1, 1)$ .

**12.134.** Find the RREF basis for the subspace of  $\mathbb{Z}_2^6$  spanned by  $v_1 = (0, 1, 1, 1, 1, 0)$ ,  $v_2 = (1, 1, 1, 0, 1, 1)$ , and  $v_3 = (1, 0, 1, 0, 0, 0)$ .

**12.135.** Let  $V$  be an  $n$ -dimensional vector space over a field  $K$ . A *flag of subspaces* of  $V$  is a chain of subspaces  $V = V_0 \supseteq V_1 \supseteq V_2 \supseteq \cdots \supseteq V_s = \{0\}$ . Suppose  $|K| = q$ . Given  $n_1, \dots, n_s$  and  $n = n_1 + \cdots + n_s$ , count the number of such flags in  $V$  such that  $\dim_K(V_{i-1}) - \dim_K(V_i) = n_i$  for  $1 \leq i \leq s$ .

**12.136.** (a) Give a linear-algebraic proof of the symmetry property  $\begin{bmatrix} n \\ k \end{bmatrix}_q = \begin{bmatrix} n \\ n-k \end{bmatrix}_q$  when  $q$  is a prime power. (b) Explain how the equality of formal polynomials  $\begin{bmatrix} n \\ k \end{bmatrix}_x = \begin{bmatrix} n \\ n-k \end{bmatrix}_x$  can be deduced from (a).

**12.137.** Let  $V$  be an  $n$ -dimensional vector space over a  $q$ -element field, and let  $X$  be the poset of all subspaces of  $V$ , ordered by inclusion. Show that the Möbius function of  $X$  is given by  $\mu_X(W, Y) = (-1)^d q^{d(d-1)/2} \chi(W \subseteq Y)$ , where  $d = \dim(Y) - \dim(W)$ . (Use 6.61.)

**12.138.** Use the recursions for  $a_n$  and  $b_n$  in §12.8 to verify the values in (12.2).

**12.139.** Give probabilistic interpretations for the rational numbers appearing as coefficients in the Maclaurin series for  $\tan x$  and  $\sec x$ .

**12.140.** Fill in the details of Step 5 of the proof in §12.8.

**12.141.** (a) List the permutations satisfying (12.3) for  $n = 1, 3, 5$ . (b) List the permutations satisfying (12.4) for  $n = 0, 2, 4$ .

**12.142.** (a) Develop ranking and unranking algorithms for up-down permutations. (b) Unrank 147 to get an up-down permutation in  $S_7$ . (c) Find the rank of 2, 5, 3, 6, 4, 8, 1, 7 among up-down permutations of length 8.

**12.143.** Let  $(q; q)_0 = 1$  and  $(q; q)_n = (1 - q)(1 - q^2) \cdots (1 - q^n)$  for  $n \geq 1$ . Consider the following  $q$ -analogues of formal trigonometric functions:

$$\sin_q = \sum_{k \geq 0} (-1)^k \frac{x^{2k+1}}{(q; q)_{2k+1}}; \quad \cos_q = \sum_{k \geq 0} (-1)^k \frac{x^{2k}}{(q; q)_{2k}} \in \mathbb{Q}(q)[[x]]; \\ \tan_q = \sin_q / \cos_q; \quad \sec_q = 1 / \cos_q.$$

Define  $q$ -tangent numbers and  $q$ -secant numbers by  $t_n = (q; q)_n \tan_q(n) \in \mathbb{Q}(q)$  and  $s_n = (q; q)_n \sec_q(n) \in \mathbb{Q}(q)$ . (a) Show that for each  $n \geq 0$ ,

$$t_n = \sum_{w \text{ satisfying (12.3)}} q^{\text{inv}(w)}; \quad s_n = \sum_{w \text{ satisfying (12.4)}} q^{\text{inv}(w)}.$$

(b) Use (a) to conclude that  $t_n, s_n \in \mathbb{N}[q]$ . Compute  $t_n$  for  $n = 1, 3, 5$  and  $s_n$  for  $n = 0, 2, 4$ .

**12.144.** Let  $t$  be the tournament with edge set

$$\{(2, 1), (1, 3), (4, 1), (1, 5), (6, 1), (2, 3), (4, 2), (5, 2), (2, 6), \\ (3, 4), (3, 5), (6, 3), (4, 5), (6, 4), (6, 5)\}.$$

Compute  $\text{wt}(t)$ ,  $\text{inv}(t)$ , and  $\text{sgn}(t)$ . Is  $t$  transitive?

**12.145.** Let  $t$  be the tournament in 12.41 and  $I$  the involution used to prove 12.46. Compute  $t' = I(t)$ , and verify directly that  $\text{wt}(t') = \text{wt}(t)$ ,  $\text{sgn}(t') = -\text{sgn}(t)$ , and  $I(t') = t$ .

**12.146.** Use induction and 9.47 to give an algebraic proof of 12.46.

**12.147.** Suppose  $x_0, x_1, \dots, x_N$  are *distinct* elements of a field  $F$ . State why the Vandermonde matrix  $[x_j^{N-i}]_{0 \leq i, j \leq N}$  is invertible. Use this to prove the fact (asserted in 2.79) that if  $p \in F[x]$  has degree at most  $N$  and satisfies  $p(x_i) = 0$  for  $0 \leq i \leq N$ , then  $p$  must be the zero polynomial.

**12.148.** A *king* in a tournament  $t$  is a vertex  $v$  from which every other vertex can be reached by following at most 2 directed edges. Show that every vertex of maximum outdegree in a tournament is a king; in particular, every tournament has a king.

**12.149.** Use the hook-length formula to compute  $f^\lambda$  for the following shapes  $\lambda$ : (a)  $(3, 2, 1)$ ; (b)  $(4, 4, 4)$ ; (c)  $(6, 3, 2, 2, 1, 1, 1)$ ; (d)  $(n, n - 1)$ ; (e)  $(a, 1^b)$ .

**12.150.** Show that  $f^{(0)} = 1$  and, for all nonzero partitions  $\lambda$ ,  $f^\lambda = \sum_{\mu} f^\mu$  where we sum over all  $\mu$  that can be obtained from  $\lambda$  by removing some corner square. Use this recursion to calculate  $f^\lambda$  for all  $\lambda$  with  $|\lambda| \leq 6$ .

**12.151.** (a) Develop ranking and unranking algorithms for standard tableaux of shape  $\lambda$  based on the recursion in 12.150. (b) Unrank 46 to get a standard tableau of shape  $(4, 3, 1)$ .

(c) Rank the standard tableau

1	3	4
2	5	8
6	7	

**12.152.** Enumerate all the hook walks for the shape  $\lambda = (4, 3, 2, 1)$  that end in the corner cell  $(2, 3)$ , and compute the probability of each walk. Use this computation to verify 12.54 in this case.

**12.153.** Suppose  $\lambda \in \text{Par}(p)$  where  $p$  is prime. (a) Show that  $p$  divides  $f^\lambda$  if  $\lambda$  is not a hook (see 10.3). (b) Compute  $f^\lambda \bmod p$  if  $\lambda$  is a hook.

**12.154.** Does the hook-length formula extend to enumerate standard tableaux of skew shape? Either adapt the probabilistic proof to this situation, or find the steps in the proof that cannot be generalized.

**12.155.** Confirm that  $\equiv_P$  and  $\equiv_K$  are equivalence relations on  $X^*$ , as asserted in §12.11.

**12.156.** Let  $T$  be the tableau in 12.62. Find an explicit chain of elementary Knuth equivalences demonstrating that  $\text{rw}(T)1 \equiv_K \text{rw}(T \leftarrow 1)$ .

**12.157.** Find the length of the longest increasing and decreasing subsequences of the word

$$w = 4135321462731132423142.$$

**12.158.** Complete the proofs of 12.64 and 12.65.

**12.159.** For any semistandard tableau  $T$ , prove that  $P(\text{rw}(T)) = T$ . Show that the set of reading words of semistandard tableaux intersects every Knuth equivalence class in exactly one point.

**12.160.** Prove 12.70 without using the RSK algorithm.

**12.161.** Show that if  $A$  is an  $N \times N$  skew-symmetric matrix with  $N$  odd, then  $\det(A) = 0$ .

**12.162.** Verify by direct calculation that  $\det(A) = (af + cd - be)^2$  for the  $4 \times 4$  matrix  $A$  in 12.72.

**12.163.** Find the Pfaffian of a general  $6 \times 6$  skew-symmetric matrix.

**12.164.** Count the number of perfect matchings for the graph shown in Figure 12.14.

**12.165.** Let  $G$  be the simple graph with  $V(G) = \{1, 2, 3, 4, 5, 6\}$  and

$$E(G) = \{\{2, 3\}, \{3, 4\}, \{4, 5\}, \{2, 5\}, \{1, 2\}, \{1, 5\}, \{3, 6\}, \{4, 6\}, \{2, 4\}\}.$$

Find all perfect matchings of  $G$ . Use this to compute  $\sum_{M \in \text{PM}(G)} \text{sgn}(M) \text{wt}(M)$ , and verify your answer by evaluating a suitable Pfaffian.

**12.166.** Compute the images of the following permutations  $w \in S_N^{\text{ev}}$  under the bijection  $S_N^{\text{ev}} \rightarrow \text{SPf}_N^2$  used in the proof of 12.83: (a)  $w = (3, 1, 5, 7)(2, 4, 8, 6)$ ; (b)  $w = (1, 4)(2, 3)(5, 7)(6, 8)$ ; (c)  $w = (2, 5, 1, 6, 8, 4, 7, 3)$ ; (d)  $w = (3, 2, 1, 5, 6, 7)(4, 8)$ .

**12.167.** Compute the images  $w$  of the following pairs  $(u, v) \in \text{SPf}_N^2$  under the bijection  $\text{SPf}_N^2 \rightarrow S_N^{ev}$  used in the proof of 12.83: (a)  $u = 13254768, v = 15283647$ ; (b)  $u = 13254768, v = 12374856$ ; (c)  $u = 15243867 = v$ . In each case, confirm that the term indexed by  $w$  in  $\det(A)$  equals the term indexed by  $(u, v)$  in  $\text{Pf}(A)^2$ .

**12.168.** Compute the exact number of domino tilings of a  $10 \times 10$  board and a  $6 \times 9$  board.

**12.169.** How many domino tilings of an  $8 \times 8$  board use: (a) 24 horizontal dominos and 8 vertical dominos; (b) 4 horizontal dominos and 28 vertical dominos?

**12.170.** Complete 12.87 by writing out  $w$  in full and showing that the placement of every new domino never causes  $\text{sgn}^*(M)$  to become negative.

**12.171.** Verify the four properties of tensor products of matrices stated just below 12.88.

**12.172.** Prove 12.89.

**12.173.** Let  $U_k$  be the matrix defined in 12.89. Show that  $\sqrt{2/(k+1)}U_k$  is a unitary matrix (i.e.,  $U^{-1} = U^*$ , where  $U^*$  is the conjugate-transpose of  $U$ ).

**12.174.** (a) Prove that, for even  $m$ ,

$$\prod_{j=1}^{m/2} 4(u^2 + \cos^2(j\pi/(m+1))) = \frac{[u + \sqrt{1+u^2}]^{m+1} - [u - \sqrt{1+u^2}]^{m+1}}{2\sqrt{1+u^2}}.$$

(b) Deduce that

$$\prod_{j=1}^{m/2} 2 \cos(j\pi/(m+1)) = 1.$$

**12.175.** Show that formula (12.12) simplifies to

$$\left\{ \begin{array}{ll} 2^{mn/2} \prod_{j=1}^{m/2} \prod_{k=1}^{n/2} \left[ x^2 \cos^2 \left( \frac{j\pi}{m+1} \right) + y^2 \cos^2 \left( \frac{k\pi}{n+1} \right) \right] & (n \text{ even}) \\ 2^{m(n-1)/2} x^{m/2} \prod_{j=1}^{m/2} \prod_{k=1}^{(n-1)/2} \left[ x^2 \cos^2 \left( \frac{j\pi}{m+1} \right) + y^2 \cos^2 \left( \frac{k\pi}{n+1} \right) \right] & (n \text{ odd}). \end{array} \right.$$


## Notes

**§12.1.** Detailed treatments of the theory of lattice paths may be found in Mohanty and Narayana [94, 98]. **§12.2.** The Chung-Feller theorem was originally proved in Chung and Feller [25]; the bijective proof given here is due to Eu, Fu, and Yeh [35]. **§12.3.** There is a growing literature on rook theory; some of the early papers in this subject are [41, 55, 56, 74]. **§12.4.** More information about parking functions may be found in [39, 43, 81, 123]. **§12.6.** Expositions of field theory may be found in Hungerford [70] or Chapter 5 of Bourbaki [19]. An encyclopedic reference for the subject of finite fields is Lidl and Niederreiter [83]. **§12.7.** For more on Gaussian elimination and RREF matrices, see linear algebra texts such as Hoffman and Kunze [69]. **§12.8.** The combinatorial interpretation of the coefficients of the tangent and secant power series is due to André [2, 4]. For more information on  $q$ -analogues of the tangent and secant series, see [6, 7, 37]. **§12.9.** Moon [97] gives a thorough account



of tournaments. The combinatorial derivation of the Vandermonde determinant is due to Gessel [52]. **§12.10.** The probabilistic proof of the hook-length formula is due to Greene, Nijenhuis, and Wilf [62]. **§12.11.** A discussion of Knuth equivalence and its connection to the RSK correspondence appears in Knuth [77]. The theorem 12.65 on disjoint monotone subsequences was proved by Greene [61]; this generalizes Schensted's original result [122] on the size of the longest increasing subsequence of a word. **§12.13.** Our treatment of the domino tiling formula closely follows the presentation in Kasteleyn's original paper [75].

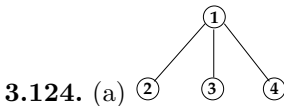
## Answers and Hints to Selected Exercises

**1.88.** (a)  $200 + 142 - 28 = 314$ ; (b)  $314 - 28 = 286$ . **1.91.**  $2^n - 2$ . **1.92.** (a)  $n(n-1)^3$ . **1.94.** about 1.8 billion years. **1.95.** (a)  $\binom{10}{3} = 120$ ; (b) 176. **1.96.**  $\binom{5}{2,3}\binom{7}{5,2} = 210$ . **1.97.** (e)  $\binom{n}{2}5^22^{1n-2}$ . **1.99.**  $n^{k/2}$  (for  $k$  even);  $n^{(k+1)/2}$  (for  $k$  odd). **1.100.** (c) uv, uw, ux, uy, vu, vw, vx, vy, wu, wv, wx, wy, xu, xv, xw, xy, yu, yv, yw, yx. **1.102.** (c) [aaa], [aab], [aac], [abb], [abc], [acc], [bbb], [bbc], [bcc], [ccc]. **1.103.** (a) (4), (3, 1), (2, 2), (2, 1, 1), (1, 3), (1, 2, 1), (1, 1, 2), (1, 1, 1, 1). **1.104.** e.g., (2, 2, 1) has picture  and word 0101. **1.107.** e.g., NNEEEN maps to NNEEEE; NEEENN maps to NEEENE; ENENNE maps to ENEEEN; etc. **1.109.** (a)  $\binom{13}{3} = 286$ ; (b)  $4^{10}$ ; (d) For (b), first compute the answer if there are only two children (cf. 1.91) or three children. This will be easier after Chapter 4. **1.110.** Each positive divisor of  $n$  has a unique prime factorization of the form  $p_1^{f_1} \cdots p_k^{f_k}$  where  $0 \leq f_j \leq e_j$  for each  $j$ . The product rule gives  $(e_1 + 1)(e_2 + 1) \cdots (e_k + 1)$  positive divisors of  $n$ . There are twice as many divisors in  $\mathbb{Z}$ . **1.111.** (c)  $\binom{N}{k} - 1$ . **1.112.** (a) e.g.,  $\Phi(12) = \{1, 5, 7, 11\}$  and  $\phi(12) = 4$ ; (b)  $\phi(p) = p - 1$ , since  $1, 2, \dots, p - 1$  must be relatively prime to  $p$ . **1.113.** (a) Use ideas from the second proof of 1.58. **1.114.** (c) Regard 'PP' as a single letter; then 1.46 gives  $\binom{10}{4,4,1,1} = 6300$ ; (d) 1.113(a) can be useful. **1.116.** (a)  $\binom{d+N-1}{d, N-1}$ . **1.118.** (a)  $6/36 = 1/6$ ; (b)  $5/36$ . **1.119.** 0.54. **1.120.** (b)  $21^5/26^5 \approx 0.344$ . **1.121.** (a)  $\binom{10}{4}/\binom{18}{4} = 7/102 \approx 0.069$ . **1.122.** (b) There are  $2^{10}$  outcomes in the sample space. We can roll zero heads in one way, or exactly one head in ten ways. So the probability of at least two heads is  $(2^{10} - 1 - 10)/2^{10} = 1013/1024 \approx 0.9893$ . **1.123.**  $\binom{10}{5,0,2,0,0,3}/6^{10} \approx 0.000042$ . **1.125.** 0.16. **1.126.** (a)  $4/5$ ; (b)  $8/15$ ; (c)  $1/2$ . **1.127.** (d) No, since  $P(A \cap B \cap D) = 0 \neq P(A)P(B)P(D)$ . **1.128.**  $100!$  has  $\lfloor 100/5 \rfloor + \lfloor 100/25 \rfloor = 24$  trailing zeroes. **1.130.** (a) 468, 559; (b) 600, 000. **1.131.** (a)  $2^{nk}$ . **1.132.**  $P(\text{full house})$  is  $(12 \cdot \binom{4}{2} + \binom{3}{2}) \cdot 12 \cdot \binom{4}{3} + 12 \cdot \binom{4}{3} \cdot 11 \cdot \binom{4}{2}) / \binom{51}{5} = 6/4165 \approx 0.00144$ . **1.133.**  $P(\text{straight})$  is  $10 \cdot 8^5 / \binom{104}{5} \approx 0.00356$  (this includes straight flushes);  $P(\text{five-of-a-kind})$  is  $13 \cdot \binom{8}{5} / \binom{104}{5} \approx 7.92 \times 10^{-6}$ . **1.135.** (a)  $3 \binom{13}{5} / \binom{39}{5} = 33/4921 \approx 0.0067$ . **1.136.**  $P(\text{straight flush} | K) = 2 / \binom{51}{4} \approx 8 \times 10^{-6}$ , so the event of getting a straight flush is not independent of  $K$ .  $P(\text{four-of-a-kind} | K) = (48 + 12) / \binom{51}{4} = 1/4165$ , so the event of getting four-of-a-kind is independent of  $K$ . **1.137.** (a)  $52!/45! = 674, 274, 182, 400$ ; (b)  $13 \cdot \binom{48}{3} \cdot 7! / |S| = 1/595 \approx 0.00168$ . **1.139.** Since  $A$  and  $B$  are nonempty, there exist  $a \in A$  and  $b \in B$ . Since  $A \neq B$ , there exists  $c$  with either  $c \in A$  and  $c \notin B$ , or  $c \in B$  and  $c \notin A$ . In the first case,  $(c, b)$  is in  $A \times B$  but not  $B \times A$ . In the second case,  $(a, c)$  is in  $A \times B$  but not  $B \times A$ . Thus,  $A \times B \neq B \times A$ . **1.141.** (b) As  $n$  tends to infinity with  $k$  held fixed, the probability that a random  $k$ -letter word using an  $n$ -letter alphabet has no repeated letters tends to 1. **1.143.** (b) Since  $A \cap B = \emptyset$ ,  $(A \cap C) \cap (B \cap C) = \emptyset$ . Thus,  $P((A \cup B) \cap C) = P((A \cap C) \cup (B \cap C)) = P(A \cap C) + P(B \cap C) = P(A)P(C) + P(B)P(C) = (P(A) + P(B))P(C) = P(A \cup B)P(C)$ . So  $A \cup B$  and  $C$  are independent. **1.144.** (c) Assume  $f$  is injective and  $g, h : W \rightarrow X$  satisfy  $f \circ g = f \circ h$ . Fix  $w \in W$ . We know  $f(g(w)) = f(h(w))$ , hence  $g(w) = h(w)$ . So  $g = h$ . For the converse, fix  $x_1, x_2 \in X$  with  $f(x_1) = f(x_2)$ . Let  $W = \{0\}$  and define  $g, h : W \rightarrow X$  by  $g(0) = x_1$ ,  $h(0) = x_2$ . Then  $f \circ g = f \circ h$  since both functions map 0 to  $f(x_1)$ . So  $g = h$  by hypothesis, which means  $x_1 = x_2$ . So  $f$  is injective. **1.146.** (a) Prove the second inequality with a counting argument. (d) Such an algorithm requires at least  $cn \log_2 n$  time steps, for some constant  $c$ . **1.147.** Each weighing has three possible outcomes (left side

heavier, right side heavier, or two sides equal). Compare to 1.146. **1.149.** Hint: compute  $f(a+1, b-1) - f(a, b)$  for  $b > 0$ . **1.150.** (d) Use the fact that for  $a, b \in \mathbb{N}^+$ ,  $\gcd(a, b) = 1$  iff there exist integers  $r, s$  with  $ar + bs = 1$ . (e) By (d),  $\phi(n) = \prod_{i=1}^k \phi(p_i^{e_i})$ . Now use 1.112(c). **1.151.** Consider the quotient and remainder when an integer  $z$  is divided by  $n$ . **1.152.** (b) Define  $g(((a, b), c)) = (a, (b, c))$  for  $a \in X, b \in Y, c \in Z$ . Note  $g$  is a bijection, since  $g'$  given by  $g'(((a, b), c)) = ((a, b), c)$  is a two-sided inverse. (c) Say  $|X| = x, |Y| = y$ , and  $|Z| = z$ . The existence of  $g$  in (b) shows  $|(X \times Y) \times Z| = |X \times (Y \times Z)|$ , so repeated use of the product rule gives  $(xy)z = x(yz)$ . **1.154.** Show that the set  $\{x \in X : x \notin f(x)\} \in \mathcal{P}(X)$  cannot be of the form  $f(z)$  for any  $z \in X$ . **1.155.** Show that each function  $f: \mathbb{N} \rightarrow \mathbb{N}[0, 1]$  cannot be surjective (cf. 1.154). **1.156.** (a) It suffices to prove  $A = X \sim g[Y \sim f[A]]$ .

**2.84.**  $BAR + BAT + BER + BET + BUR + BUT + CAR + CAT + CER + CET + CUR + CUT + HAR + HAT + HER + HET + HUR + HUT$ . **2.86.**  $\binom{9}{2,3,1,3} = 5040$ . **2.88.**  $-160$ . **2.90.** For the first identity, expand  $(-1+1)^n$  using the binomial theorem. For a combinatorial proof, first move negative terms (indexed by odd  $k$ ) to the other side of the equation. **2.92.**  $m^n/n!$ . **2.93.** e.g.,  $\binom{9}{2} = 8+28 = 36$ ;  $\binom{9}{3} = 28+56 = 84$ ; so  $\binom{10}{3} = \binom{9}{2} + \binom{9}{3} = 120$ . **2.95.** e.g.,  $T(2, 7) = 27, T(3, 7) = 27+48 = 75$ . **2.97.** 5524 (use a recursion). **2.98.** 136 (draw a picture). **2.100.** Find a bijection between these partitions and Dyck paths. **2.102.** Integer partitions of 8 into 3 parts are  $(6, 1, 1), (5, 2, 1), (4, 3, 1), (4, 2, 2)$ , and  $(3, 3, 2)$ , so  $p(8, 3) = 5$ . Using the recursion,  $p(8, 3) = p(7, 2) + p(5, 3) = p(6, 1) + p(5, 2) + p(4, 2) + p(2, 3) = \dots = 5$ . **2.103.**  $p(13) = p(12) + p(11) - p(8) - p(6) + p(1) = 101, p(14) = 135$ . **2.105.** e.g.,  $S(9, 2) = 255, S(9, 3) = 3025, S(9, 4) = 7770, S(10, 9) = 45$ . **2.106.**  $B(9) = 21, 147, B(10) = 115, 975$ . **2.107.** e.g.,  $s(8, 1) = -5040, s(8, 2) = 13, 068, s(8, 3) = -13, 132$ . **2.110.** (a)  $2^{n^2}$ ; (b)  $2^{n^2-n}$ ; (c)  $2^{n^2-n}$ . **2.113.** (b)  $\{\{1, 7\}, \{2, 4, 6\}, \{3\}, \{5\}\}$ . **2.115.** Use a recursion. **2.119.** (a) Mapping  $w_1 w_2 \dots w_n$  to its reversal  $w_n \dots w_2 w_1$  defines a bijection from  $S_n^{231}$  to  $S_n^{132}$ . (c) Verify that the numbers  $|S_n^{312}|$  satisfy the Catalan recursion by classifying  $w \in S_n^{312}$  based on the index  $k$  for which  $w_k = 1$ . Such a  $w$  can be written  $w = w'1w''$ , where  $w'$  is any 312-avoiding permutation of  $\{2, 3, \dots, k\}$ , and  $w''$  is any 312-avoiding permutation of  $\{k+1, \dots, n\}$ . **2.120.** (a) Given  $(g_0, \dots, g_{n-1}) \in G_n$ , replace  $g_0$  by an N and, for  $1 \leq i < n$ , replace  $g_i$  by  $g_{i-1} + 1 - g_i$  E's followed by an N. Add  $g_{n-1} + 1$  E's at the end. Check that this gives a bijection from  $G_n$  to the set of Dyck paths of order  $n$ . **2.121.** (a) NNNEENNEENEENNENEENEE; (c) 3 2 5 4 6 1 8 9 7 11 10. **2.124.** (a) To build a set partition of  $\{1, 2, \dots, n\}$  into 2 blocks, choose any subset  $U$  of  $\{1, 2, \dots, n-1\}$  except  $\emptyset$  to be one block, and let the other block be  $\{1, 2, \dots, n\} \sim U$ . So  $S(n, 2) = 2^{n-1} - 1$ . (b) A surjection from an  $n$ -element set to an  $n$ -element set is automatically a bijection, by 1.27. So  $\text{Surj}(n, n) = n!$  by 1.28. **2.127.** Find a recursion for  $f(n, a, b)$ , the number of paths of length  $n$  from  $(0, 0)$  to  $(a, b)$ . Use a computer to obtain  $f(10, 0, 0)$ . **2.129.** (a)  $s'(\mu, k) = \sum_{1 \leq i_1 < i_2 < \dots < i_{n-k} \leq n} \mu_{i_1} \mu_{i_2} \dots \mu_{i_{n-k}}$ ; (c)  $s'(\mu, 0) = s'(\mu, 1) = 0, s'(\mu, 2) = 360, s'(\mu, 3) = 717$ , etc. **2.130.** (b) Let  $g_n$  be the number of subsets of  $\{1, 2, \dots, n\}$  that do not contain two consecutive integers. Check that  $g_0 = 1 = f_0$  and  $g_1 = 2 = f_1$ . To see  $g_n = g_{n-1} + g_{n-2}$  for  $n \geq 2$ , consider whether  $n$  does not or does belong to the subset counted by  $g_n$ ; in the latter case,  $n-1$  cannot also appear in the subset. Induction now shows  $g_n = f_n$  for  $n \geq 0$ . **2.131.** (a) Note  $f_{2n} = f_{2n-1} + f_{2n-2} = (f_{2n-2} + f_{2n-3}) + f_{2n-2}$ . Now use  $f_{2n-2} = f_{2n-3} + f_{2n-4}$  to eliminate  $f_{2n-3}$ . Initial conditions are  $a_0 = 1$  and  $a_1 = 3$ . **2.133.** (b) Divide by  $x-1$  and let  $n$  tend to infinity in (a) to get  $\sum_{m=0}^{\infty} x^m = 1/(1-x)$  for  $|x| < 1$ . **2.134.** (a) Use induction, noting that  $\phi$  and  $\psi$  satisfy the equation  $x^2 = x + 1$ . (b) For an algebraic proof, use (a) and 2.133. **2.135.** Hint: For any function  $f: X \rightarrow Y$ ,  $\{(a, b) \in X \times X : f(a) = f(b)\}$  is an equivalence relation on  $X$ . **2.136.** Given a path  $\pi$  counted by  $C_{n,k}$ , choose  $r$  so that  $\pi$  arrives at  $(n-k-r, n-k)$  by a north step. **2.137.** One approach to the combinatorial proof is to use 12.1, noting that  $\gcd(k, p-k)$  must be 1. **2.138.** This method of proof is due to Leibniz. **2.139.** (c) 1, 1, 3, 13, 75, 541. **2.140.**

(a)  $B_1(0) = 1$ ,  $B_1(1) = 0$ ,  $B_1(n) = \sum_{k=1}^{n-1} \binom{n-1}{k} B_1(n-1-k)$ ; so  $B_1(2) = 1$ ,  $B_1(3) = 1$ ,  $B_1(4) = 4$ ,  $B_1(5) = 11$ ,  $B_1(6) = 41$ . **2.141.**  $p_d(n, k) = 0$  for  $n < 0$  or  $k < 0$  or  $k > n$ ;  $p_d(0, k) = \chi(k = 0)$ ;  $p_d(1, k) = \chi(k = 1)$ ;  $p_d(n, k) = p_d(n-1, k-1) + p_d(n-k, k-1)$  for  $n > 1$ ,  $1 \leq k \leq n$ . **2.145.** (a) Fix  $v \in W$ . Write  $B = (v_1, \dots, v_n)$ ,  $C = (w_1, \dots, w_n)$ ,  $[v]_B = (s_1, \dots, s_n)^T$ , and  $[v]_C = (r_1, \dots, r_n)^T$ . On one hand,  $v = \sum_i r_i w_i$ . On the other hand,  $v = \sum_j s_j v_j = \sum_j s_j \sum_i a_{ij} w_i = \sum_j \sum_i s_j a_{ij} w_i = \sum_i \left( \sum_j a_{ij} s_j \right) w_i$ . Equating the coefficients of  $w_i$  gives  $r_i = \sum_j a_{ij} s_j$  for all  $i$ , which says  $[v]_C = A[v]_B$ . **2.150.** (a) To prove identities involving  $\oplus$  or  $\otimes$ , these facts can be helpful: (i)  $c \bmod n = c - qn$  for some  $q \in \mathbb{Z}$ ; (ii) for  $u, v \in \mathbb{Z}_n$ ,  $u = v$  if  $n$  divides  $u - v$ . For associativity of  $\otimes$ , note  $(a \otimes b) \otimes c = (ab - qn) \otimes c = (ab - qn)c - rn = (ab)c + n(-qc - r)$  for some  $q, r \in \mathbb{Z}$ . Similarly,  $a \otimes (b \otimes c) = a(bc) + n(-sa - t)$  for some  $s, t \in \mathbb{Z}$ . As  $(a \otimes b) \otimes c \in \mathbb{Z}_n$  and  $a \otimes (b \otimes c) \in \mathbb{Z}_n$  and their difference is divisible by  $n$ , these elements are equal. **2.151.**  $|M_n(R)| = |R|^{n^2}$ . **2.153.** (a) To prove the right distributive law in  $R$ , fix  $f, g, h \in R$  and  $n \in \mathbb{Z}$ . Check that both  $(f \oplus g) \circ h$  and  $(f \circ h) \oplus (g \circ h)$  send  $n$  to  $f(h(n)) + g(h(n))$ . (b) Be sure to check that  $S$  is closed under  $\oplus$  and  $\circ$ . **2.154.** In the induction step, reindex the summations and use 2.25. **2.156.** (b)  $D^n(f_1 f_2 \cdots f_s) = \sum_{n_1 + \cdots + n_s = n} \binom{n}{n_1, \dots, n_s} D^{n_1}(f_1) D^{n_2}(f_2) \cdots D^{n_s}(f_s)$ . **2.157.** (c) The degree 2 polynomial  $x^2 - 1 \in \mathbb{Z}_8[x]$  has four roots in  $\mathbb{Z}_8$  (1, 3, 5, and 7). **2.158.** Use induction on  $n$  and 2.157(b). You may need 7.44 also. **2.159.** (b) Use 2.158. **2.160.**  $|A_n| = C_n$  (see §12.2). **2.161.** see [22].



**3.124.** (a) (e) **3.125.** (b) Choose to include or exclude each of the  $n^2$  ordered pairs  $(v, w)$  in the edge set of the simple digraph. By the product rule, the answer is  $2^{n^2}$ . **3.127.** The bijection maps  $(V, E)$  to the relation  $R \subseteq V \times V$  such that for  $u, v \in V$ ,  $(u, v) \in R$  iff  $\{u, v\} \in E$ . **3.129.** There are eleven

isomorphism classes for four-vertex simple graphs. **3.130.** (a)  $\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ ; (g) an  $n \times n$

matrix with zeroes on the main diagonal and ones elsewhere. **3.131.** For  $1 \leq k \leq 5$ , answers are 1, 1, 7, 9, and 57. **3.132.** There are 64 walks from 0 to 0, 36 from 0 to 1, 120 from 0 to 2, 76 from 1 to 2, etc. **3.134.** (a) Use induction on the length of the walk. (b) No; consider a graph with a single edge. **3.135.** (a) if  $i = j$ , zero; if  $i \neq j$ ,  $\sum_{k \neq i, j} A(i, k)A(k, j) = A^2(i, j) - A(i, j)(A(i, i) + A(j, j))$ . **3.136.** (b) 8. **3.138.** (a) 8 paths; (b) 13 paths. **3.139.**  $2^{n-2}$  for  $n \geq 2$ . **3.140.**  $\sum_{k=1}^n k!S(n, k)$ , the number of ordered set partitions of  $\{1, 2, \dots, n\}$  (see 2.139). **3.142.** (a)  $\deg(G) = [2, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 6]$ ; sum of degrees is  $40 = 2 \cdot 20 = 2|E|$ . **3.144.** (a)  $\deg(C_n) = [2, 2, \dots, 2]$  where 2 occurs  $n$  times; (c)  $(n-1)!/2$ ; (d)  $p(n)$ , the number of integer partitions of  $n$ . **3.146.** The smallest example has five vertices. **3.147.** The statement is false. **3.148.** Find a three-vertex example. **3.149.** Note  $\sum_{e \in E} M(v, e) = \deg_G(v)$  for  $v \in V$  and  $\sum_{v \in V} M(v, e) = 2$  for  $e \in E$ . **3.150.** (b) Digraph has edges  $(0, 1)$ ,  $(1, 2)$ ,  $(2, 5)$ ,  $(3, 3)$ ,  $(4, 3)$ ,  $(5, 5)$ ,  $(6, 2)$ .  $C = \{3, 5\}$ ,  $S_3 = \{3, 4\}$ , and  $S_5 = \{0, 1, 2, 5, 6\}$ . **3.152.** Study the walk through the functional digraph of  $f$  determined by the sequence  $(x_i : i \geq 0)$ . **3.153.** (a) With the notation of 3.152, the algorithm computes  $\gcd(x_{2i} - x_i, N)$  for  $i = 1, 2, \dots$  until this gcd is not 1. For some  $i > 0$ ,  $x_{2i} = x_i$ , so the gcd for this  $i$  is  $N$ , and the algorithm terminates. However, the algorithm may terminate sooner if  $\gcd(v - u, N)$  is a proper divisor of  $N$ . (b) For  $N = 77$ : in step 1,  $(u, v, d) = (1, 2, 1)$ . In step 2,  $(u, v, d)$  is  $(2, 26, 1)$ , then  $(5, 26, 7)$ . So 7 is a divisor. **3.154.** (a)  $(k-1)(k-2) \cdots (k-s)/k^s$ . (b) Observe that, if  $x_0, \dots, x_{i-1}$  are all distinct and  $i \geq \sqrt{k}$ , then the probability that  $S = i$  is at least  $1/\sqrt{k}$ . (c)  $N$  has a prime divisor  $p \leq \sqrt{N}$ . Consider  $Y = \{0, 1, \dots, p-1\}$  and  $g : Y \rightarrow Y$  given by  $g(y) = (y^2 + 1) \bmod p$ . Assume

$g$  behaves approximately like a random function. **3.156.** (a) 1960. **3.158.**  $n! / \prod_i (a_i! i^{a_i})$ . **3.161.** The function sends  $1, 2, \dots, 17$  to  $1, 12, 3, 4, 10, 17, 15, 8, 3, 3, 12, 1, 4, 10, 1, 4, 17$ , respectively. **3.163.**  $(1, a, 3, f, 5, m, 2, k, 4)$ . **3.164.** Each cycle is a strong component; and each  $v$  not on a cycle is in a component by itself. **3.168.** 38. **3.170.** (a) A walk of odd length in  $G$  that starts in  $A$  (resp.  $B$ ) must end in  $B$  (resp.  $A$ ), so could not be a cycle. **3.171.**  $2^{mn}$ . **3.172.** (a)  $kn/2$  (by 3.34); (b) argue that  $G$  has  $k|A|$  edges and also  $k|B|$  edges. **3.173.** The statement is false. **3.175.**  $2(n-k)$ . **3.179.** The case where  $|V(T_i)| = 1$  for some  $i$  can be handled directly. Otherwise, use induction and pruning. **3.180.** Use induction on  $m$  and pruning. (The result is due to Smolenskii.) **3.181.**  $1, 1, 1, 2, 3, 6, 11$ . **3.182.** (a)  $p(n, 3)$ . **3.183.**  $n!S(n-2, n-k)/k!$  (see [112]). **3.184.** (b)  $(n-2)n^{n-3}$ . **3.186.** In an  $n$ -vertex tree, delete the unique edge incident to the largest vertex that leads towards the smallest vertex. **3.187.** (a) 2274547; (b) edges are  $\{8, 1\}, \{6, 3\}, \{4, 5\}, \{2, 5\}, \{5, 1\}, \{1, 7\}, \{7, 3\}, \{3, 0\}$ . **3.190.** (a) 8580. **3.192.** Use 3.91 with  $n = 1, k_0 = m, k_1 = 0, k_2 = m-1, s = 2m-1$  to get  $\frac{1}{2m-1} \binom{2m-1}{m, m-1} = C_{m-1}$  parenthesizations. **3.194.** (a)  $x(x-1)^3$ ; (c)  $x(x-1)(x-2)^2$ . **3.197.** (a) 4; (b) 108,000. **3.198.** (a)  $\sum_{k=1}^n \binom{n}{k} 2^{k(k-1)/2}$ ; (b)  $2^n - 1$ . **3.203.** Let  $X$  be the set of pairs  $(T, e)$  where  $T$  is a spanning tree of  $G_{n-1}$  and  $e$  is one of the  $k$  edges of the rightmost  $k$ -gon in  $G_n$ . Map  $X$  bijectively to the set of trees that span  $G_{n-2}$  or  $G_n$ . Initial conditions are  $\tau(G_0) = 1$  and  $\tau(G_1) = k$ . **3.204.** (a) Let  $X$  be the set of pairs  $(e, v)$  where  $e \in E(G)$  and  $v \in V(G)$  is not an endpoint of  $e$ . Choose  $e$  and then  $v$  to see  $|X| = |E(G)|(n-2)$ . Choose  $v$  and then  $e$  to see  $|X| = \sum_{v \in V(G)} |E(G \sim v)|$ . **3.205.** (a) 1 tree (the whole graph), and  $\det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 1$ . (c) 8 trees, and  $\det \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix} = 8$ . **3.208.** The determinant is the characteristic polynomial of  $uJ$ . What are the minimal polynomial and eigenvalues (with multiplicities) of  $uJ$ ? **3.209.** Use 3.208 with  $m = n-1, t = n$ , and  $u = 1$ . **3.210.** 3.208 can be useful here. **3.212.** (a)  $G$  is connected and every vertex has even degree. **3.214.** Construct a digraph with vertex set  $A^{k-1}$  and labeled edges  $y_1 \cdots y_{k-1} \xrightarrow{z} y_2 \cdots y_{k-1} z$  for each  $z \in A$ . Argue that this digraph has a closed Eulerian tour, and use this to construct the word  $w$ . **3.215.** (a) 10 vertices, 15 edges; (d) No, since  $G$  has cycles of odd length. **3.217.** Use the product rule and 3.215(c),(e) to show each  $e$  belongs to four 5-cycles. So  $G$  has  $15 \cdot 4/5 = 12$  5-cycles. **3.218.** (a) Define a bijection between the set of such cycles and  $V(G)$ . In the second picture of  $G$  shown in 3.215(b), the outer cycle  $(A, B, C, D, E, F, A)$  maps to the central vertex  $J$ . (b) 10.

**4.60.** (a)  $|S \cup T| = 15 + 13 - 6 = 22$ ; (b)  $|S \cup T \cup U| = 15 + 13 + 12 - 6 - 3 - 4 + 1 = 28$ ; (c) 17. **4.62.** 2143, 2341, 2413, 3142, 3412, 3421, 4123, 4312, 4321. **4.63.** (d)  $d_{10} = 10d_9 + 1 = 1, 334, 961$ . **4.64.** (b)  $\phi(11) = 10, \mu(11) = -1, \tau(11) = 2, \sigma(11) = 12$ . (c)  $\phi(28) = 12, \mu(28) = 0, \tau(28) = 6, \sigma(28) = 56$ . **4.65.** (a)  $1 + 1 + 2 + 2 + 2 + 4 + 4 + 8 = 24$  and  $24 - 12 - 8 + 4 = 8 = \phi(24)$ . **4.67.**  $\binom{52}{5} - 4\binom{39}{5} + 6\binom{26}{5} - 4\binom{13}{5} = 685,464$ . **4.71.**  $m(n-1)!S(m-1, n-1)$ . **4.72.** (a)  $x^4 - 4x^3 + 6x^2 - 3x$ . **4.74.** (a) If  $m$  has prime factorization  $\prod_i p_i^{e_i}$ , then  $\sigma_k(m) = \prod_i (1 - p_i^{k(e_i+1)}) / (1 - p_i^k)$ . (b) Use 4.30. **4.76.** Writing permutations in one-line form, 1234 pairs with 2134, 1243 pairs with 2143, 1324 pairs with 2314, ..., and 4321 pairs with 4312. **4.78.**  $w$  is a 6-cycle  $(1, 4, 2, 3, 6, 5)$ . If we splice 7 in just before the 2 (say), we get the 7-cycle  $(1, 4, 7, 2, 3, 6, 5)$ , which is  $4367152 \in D_7$ . If we pair 8 with 5 (say) in a 2-cycle and relabel  $w$  to be  $(1, 4, 2, 3, 7, 6)$  we get  $43728165 \in D_8$ . **4.82.** (a) Expand  $(2-1)^n = 1$  using the binomial theorem. (b) Consider  $n$ -letter words using letters  $\{a, b, c\}$ , with an appropriate definition of signs. **4.83.**  $(a-b)^n$ . **4.85.** (a) In the case  $S \neq T$ , define an involution based on whether a fixed element of  $T \sim S$  is or is not in  $U$ . **4.87.** (a)  $(3^n - 1)^n$ ; (b)  $(3^n - 2^n)^n$ ; (c) use inclusion-exclusion. **4.88.**  $\sum_{k=0}^n (-1)^k \binom{n}{k} 2^{(n-k)^2}$ . **4.92.** (a) In 4.11, let  $S_i$  be the set of words in  $\mathcal{R}(1^3 2^3 \cdots n^3)$  that contain three  $i$ 's in a row. To compute (say)  $|S_1 \cap S_2|$ , think of 111 (and 222) as a single letter. **4.95.** (a) Use 4.48, classifying signed

chains from  $x$  to  $z$  based on the next-to-last point  $y$ . **4.96.** For  $x, y \in X$ , define  $x < y$  iff there exist  $k > 0$  and  $x_0, \dots, x_k \in X$  such that  $x_0 = x$ ,  $x_k = y$ , and  $(x_{i-1}, x_i) \in R$  for

$$1 \leq i \leq k. \quad \mathbf{4.97.} \quad (\text{a}) \quad Z = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad M = Z^{-1} = \begin{bmatrix} 1 & -1 & -1 & -1 & 2 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (\text{b}) \quad \text{For}$$

instance,  $\mu(a, e) = 2$  since there are three positive chains of length 2 (through  $b, c$ , or  $d$ ) and one negative chain  $(a, e)$  of length 1. **4.99.** Use 4.95 and induction on the number of elements in the interval. **4.100.**  $\mu_X(a, b) = \mu_{X_1}(a, b)$  if  $a, b \in X_1$ ;  $\mu_X(a, b) = \mu_{X_2}(a, b)$  if  $a, b \in X_2$ ; and  $\mu_X(a, b) = 0$  otherwise. **4.102.** One can either invert  $Z$  using the algebra of block matrices, or count signed chains in  $X$ . **4.105.** (b) Repeated use of 4.104 shows that the events  $X \sim S_1, \dots, X \sim S_n$  are independent. Since  $P(X \sim S_i) = 1 - p_i$ , the desired probability is  $\prod_{i=1}^n (1 - p_i)$ . **4.106.** The left side counts  $i$ -element subsets of  $\{1, 2, \dots, n\}$  in which some elements in the subset have been marked with negative signs. If  $i > 0$ , take the least element in the subset and flip its status (marked or not). **4.109.**  $n!$ . **4.111.**  $A^{-1}(i, j) = (-1)^{i+j} \binom{i-1}{j-1}$ . **4.113.** 41, 304. **4.115.** Apply 4.11, letting  $S_i$  be the set of compositions  $(a_1, \dots, a_k)$  with  $a_1 + \dots + a_k = n$  and  $a_i > m$ . **4.116.** One might try inclusion-exclusion, but it may be easier to use 2.140. For  $m = 11$  and  $n = 4$ , the answer is 1,367,520. **4.117.** Try to generalize one of the proofs in §4.3. **4.118.** (a)  $\sum_{i=0}^{n-k} (-1)^i \frac{n!}{k!i!}$ . (b) combinatorial proof: choose  $k$  of the  $n$  objects to be fixed points ( $\binom{n}{k}$  ways); then choose a derangement of the remaining  $n - k$  objects ( $d_{n-k}$  ways); now use the product rule. **4.119.** 616. **4.120.** For  $n > 0$ , the sum is  $(-1)^n F_{n-1} - 1$  (as can be proved by induction). **4.122.** 1. **4.124.**  $(-1)^{n-1} (n-1)!$ . **4.126.** Starting at  $x_0 \in \text{Fix}(I)$ , repeatedly apply  $f$ , then  $J$ , then  $f^{-1}$ , then  $I$ , until some application of  $f$  leads to an element in  $\text{Fix}(J)$ . Argue that this process must terminate and is a bijection. **4.127.** Starting at  $x \in B$ , keep applying  $g$  until  $C$  is reached. **4.128.** See [28] for one approach. **4.129.** See, e.g., [87], which provides an automatic method for converting a bijective proof of a matrix identity  $AB = I$  into a bijective proof of  $BA = I$ . **4.130.** Careful analysis of the proofs of 4.24 and 4.25 leads to a recursively defined map. For other approaches, see [110, 138]. **4.131.** (d) Use (c) and 4.95. The key fact, which can be proved by induction, is that  $\mu_{X_n}(\{\{1\}, \{2\}, \dots, \{n\}\}, \{\{1, 2, \dots, n\}\}) = (-1)^{n-1} (n-1)!$ . See [127], Section 3.10, for a full discussion.

**5.45.**  $g+f : \{a, b, c, d, e\} \rightarrow \underline{5}$  sends  $a$  to 3,  $b$  to 4,  $c$  to 2,  $d$  to 1, and  $e$  to 0. **5.46.** (a) 23; (c) 31; (e) (4, 2). **5.47.** (b) (1, 1, 1, 0, 1); (c) 153. **5.48.** (a) 91; (d) (2, 0, 1, 0). **5.49.** (b)  $g_0 + g_1 + g_2 + g_3$  maps (0, 0) to 0, (1, 0) to 1, (2, 0) to 2, (0, 1) to 3, ..., (1, 3) to 10, and (2, 3) to 11. **5.50.** (a) 7944; (c) 299,140. **5.51.** (a) good; (d) this. **5.53.** (a) 12,129; (f) BNA. **5.54.** (a) 11,553; (f) GNA. **5.55.** *Suggestion:* use a space character. (b) 1875. **5.56.** (a) Choose the digits from left to right to get  $8 \cdot 9 \cdot 9 \cdot 4 = 2592$ ; (c) rank of 2500 is 504; (d) 2501 unranks to 9742. **5.57.** (a) rank of LEVEL is 7561; (b) 12,662 unranks to STATS. **5.58.** (a) 6,187,926. **5.60.** (a) rank of bfdc is  $p_{6,5,4,3}(1, 4, 2, 1) = 115$ . **5.61.** (a) rank of 42153 is  $p_{5,4,3,2,1}(3, 1, 0, 1, 0) = 79$ . (b)  $p_{5,4,3,2,1}^{-1}(46) = (1, 3, 2, 0, 0)$ , so 46 unranks to the permutation 25413. **5.63.** (a) rank of  $\{b, c, d, h\}$  is 38; (b) 40 unranks to  $\{a, c, e, h\}$ . **5.65.** (a) rank of bbccacba is 253; (b) 206 unranks to bbabccac. **5.67.** (a)  $r_{9,3}((3, 3, 3)) = 0$ ; (b)  $u_{12,3}(6) = (3, 3, 1, 1, 1, 1, 1)$ . **5.68.** (4, 4), (4, 3, 1), (4, 2, 2), (4, 2, 1, 1), (4, 1, 1, 1, 1). **5.71.** (a) rank of  $\{\{1, 3\}, \{2, 4, 5\}\}$  is 13; (b) 247 unranks to  $\{\{1, 6\}, \{2, 3\}, \{4, 7\}, \{5\}\}$ . **5.73.** (a) This hand is generated by the data  $(x, B, y, C) = (3, \{\clubsuit, \diamond, \heartsuit\}, 9, \{\diamond, \spadesuit\})$ , so the rank is  $p_{13,4,12,6}(2, 0, 7, 4) = 622$ . (b)  $p_{13,4,12,6}^{-1}(3082) = (10, 2, 9, 4)$ , giving data  $(x, B, y, C) = (J, \{\clubsuit, \heartsuit, \spadesuit\}, 10, \{\diamond, \spadesuit\})$ . The hand is  $\{J\clubsuit, J\heartsuit, J\spadesuit, 10\diamond, 10\spadesuit\}$ . **5.77.** (a) NENENENE, NENENNEE, NENNEENE, NENNE-NEE, NENNEEEE, NNEENENE, NNEENNEE, NNENEENE, NNNEEENE, NNENE-NEE, NNENNEEEE, NNNEENEE, NNNENEEE, NNNNEEEE. **5.79.** (a)  $T$  maps to the word  $0005654100 \in \underline{11}^9$ , which has rank 900,350. **5.80.** (a) ccbacbd; (c) 01101101. **5.81.**

(a) ccbabcd; (c) 01100111. **5.82.** First successor is NNNENEENNEENNNEEEE; first predecessor is NNNENEENNEENNEENEE. **5.84.** For  $b < 0$ , apply 5.3 to  $a$  and  $|b|$ , and then replace  $q$  by  $-q$ . **5.85.** Obtain an initial  $q$  and  $r$  using 5.84. If  $r$  is too big, modify it by adjusting  $q$  appropriately. **5.87.** (b) Take  $F = \mathbb{Z}$ ,  $f = 3x$ ,  $g = 2x$ . If  $f = qg + r$  as in (a), then  $r$  and  $q$  must be integers and hence  $2q = 3$ , which is impossible. (c) Yes, since the proof of (a) works if the leading coefficient of  $g$  has a multiplicative inverse in  $F$ . **5.89.** (a) ( $\Leftarrow$ ): Assume  $\gcd(s, t) = 1$ , so  $\text{lcm}(s, t) = st$ . We show  $f$  is one-to-one. Let  $x, y \in \underline{st}$  with  $f(x) = f(y)$ . Then  $x \bmod s = y \bmod s$  and  $x \bmod t = y \bmod t$ , so  $s$  divides  $x - y$  and  $t$  divides  $x - y$ . Thus,  $x - y$  is a common multiple of  $s$  and  $t$  strictly between  $-st$  and  $st$ . Since  $\text{lcm}(s, t) = st$ ,  $x - y = 0$ , so  $x = y$ . Since  $f$  is one-to-one, the image of  $f$  has size  $|\underline{st}| = st = |\underline{s} \times \underline{t}|$ . So  $f$  is a bijection. **5.91.** Develop a recursive algorithm using the base-2 expansion of  $e$  and the identities  $x^{2^f} \bmod n = (x^f \bmod n)^2 \bmod n$  and  $x^{2^{f+1}} \bmod n = ((x^{2^f} \bmod n) \cdot x) \bmod n$ . **5.93.** (b) Rename  $a, b, c, d, e$  to be  $0, 1, 2, 3, 4$ . Using the bijection in the second proof in §1.11,  $[b, b, c, d, d, d]$  maps to the subset  $\{1, 2, 4, 6, 7, 8\}$  of  $\{0, 1, \dots, 9\}$ , which has rank 70. 132 unranks to  $\{1, 2, 5, 6, 7, 9\}$ , which maps to the multiset  $[b, b, d, d, d, e]$ . **5.94.** Let  $n$  tend to infinity in 5.22. **5.97.** One approach is to use a bijection to map partitions of  $n$  with  $k$  distinct parts to partitions of  $n - \binom{k}{2}$  with first part  $k$ , and then apply the algorithms in §5.8. **5.100.** As seen in §1.13, a three-of-a-kind hand is uniquely determined from data  $(x, B, C, s_1, s_2)$ , where  $x \in \text{Values}$ ,  $B$  is a three-element subset of Suits,  $C$  is a two-element subset of Values  $\sim \{x\}$ , and  $s_1, s_2 \in \text{Suits}$ . Let the rank of the hand be  $p_{13,4,66,4,4}(r(x), r(B), r_x(C), r(s_1), r(s_2))$ . **5.102.** (a) The straight is determined from the lowest value  $v$  (which is in  $\{A, 2, \dots, 10\}$ ) and a word in Suits<sup>5</sup>. The given hand has  $v = 4$  and word  $\heartsuit\heartsuit\clubsuit\diamondsuit$ , leading to a rank of  $p_{10,4,4,4,4,4}(3, 2, 2, 0, 1, 1) = 3717$ .  $p_{10,4,4,4,4,4}^{-1}(1574) = (1, 2, 0, 2, 1, 2)$ , leading to the hand  $\{2\heartsuit, 3\clubsuit, 4\heartsuit, 5\diamondsuit, 6\heartsuit\}$ . **5.105.** *Ranking:* If  $n \notin S$ , let  $r_n(S) = r_{n-1}(S)$ . If  $n \in S$ , let  $r_n(S) = f_{n-1} + r_{n-2}(S \sim \{n\})$ . *Unranking:* If  $0 \leq k < f_{n-1}$ , let  $r_n^{-1}(k) = r_{n-1}^{-1}(k)$ . Otherwise, let  $r_n^{-1}(k) = r_{n-2}^{-1}(k - f_{n-1}) \cup \{n\}$ . **5.106.** (b) 412. **5.107.** Successor of  $(9, 3)$  is  $(10, 2)$ . **5.108.** Successor of  $(7, 4, 2, 1)$  is  $(7, 4, 3)$ . **5.110.** (b) For one ordering (based on 5.30), the successor is  $\{J\clubsuit, J\diamondsuit, J\spadesuit, 9\diamondsuit, 9\heartsuit\}$ . (c)  $\{J\clubsuit, J\diamondsuit, J\spadesuit, 9\clubsuit, 9\diamondsuit\}$ . **5.113.** Let  $n = n_1 + \dots + n_k$ . To choose  $w = w_1 \dots w_n$ , randomly choose  $w_1$  to be  $a_i$  with probability  $n_i/n$  ( $1 \leq i \leq k$ ). Decrement  $n_i$  by 1, and recursively choose  $w_2 \dots w_n$ . **5.117.** No. For instance, when  $n = 3$ , check that the probability of generating  $w = 123$  is  $4/27 \neq 1/6$ . **5.118.** Yes (use induction on  $i$ ). **5.120.** Find a recursion based on the initial letter of the word, and convert this to a ranking algorithm.

**6.50.** (a) e.g.,  $\text{inv}(1432) = 3$ ,  $\text{des}(1432) = 2$ ,  $\text{maj}(1432) = 5$ ;  $\text{inv}(3124) = 2$ ,  $\text{des}(3124) = 1$ ,  $\text{maj}(3124) = 1$ . (b)  $G_{S_4, \text{inv}}(x) = G_{S_4, \text{maj}}(x) = 1 + 3x + 5x^2 + 6x^3 + 5x^4 + 3x^5 + x^6$ ;  $G_{S_4, \text{des}}(x) = 1 + 11x + 11x^2 + x^3$ . **6.51.** For  $w = 314423313$ ,  $\text{inv}(w) = 16$ ,  $\text{Des}(w) = \{1, 4, 7\}$ ,  $\text{des}(w) = 3$ , and  $\text{maj}(w) = 12$ . **6.52.** e.g., the path NNENEENE (or 00101101) has area 2 and major index 9. **6.54.** Build  $w$  in  $S$  by choosing  $w_i \in \underline{n}$  for  $i = 1, 2, \dots, k$ . The generating function for the  $i$ th choice is  $[n]_x$ , so the product rule for weighted sets gives  $G_{S, \text{wt}}(x) = [n]_x^k$ . **6.56.**  $(1+x)^n = [2]_x^n$ . **6.57.** One approach is to use a bijection between multisets and lattice paths. **6.58.** (a) 3017. **6.59.** (a) 139. **6.61.** To begin, introduce suitable signs and weights on the set of all subsets of  $\underline{n}$ . **6.62.** (c)  $1 + x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + 6x^6 + 6x^7 + 6x^8 + 6x^9 + 5x^{10} + 4x^{11} + 3x^{12} + 2x^{13} + x^{14} + x^{15}$ ; (e)  $1 + x^2 + x^3 + x^4 + x^5 + x^6 + x^8$ . **6.63.** (a)  $[6]_x = (1+x)(1+x+x^2)(1-x+x^2)$  in  $\mathbb{Z}[x]$ ; (b)  $[6]_x = (x - [1+i\sqrt{3}]/2)(x - [-1+i\sqrt{3}]/2)(x+1)(x - [-1-i\sqrt{3}]/2)(x - [1-i\sqrt{3}]/2)$  in  $\mathbb{C}[x]$ . **6.64.** (f) The partitions are  $(0)$ ,  $(1)$ ,  $(2)$ ,  $(1, 1)$ ,  $(2, 1)$ , and  $(2, 2)$ , so  $\left[\frac{4}{2}\right]_x = 1 + x + 2x^2 + x^3 + x^4$ . **6.65.** Compare to the second proof of 1.46. **6.66.** (c) For all  $z \in T_1$ ,  $w_3((g \circ f)(z)) = w_3(g(f(z))) = w_2(f(z)) = w_1(z)$ . **6.68.** (b) 87654321; weight is  $6 + 6 + 16 = 28$ . **6.69.** (c) (45321, 4321, 111001000); weight is  $9 + 6 + 18 = 33$ . **6.71.** Generalize the second proof of 6.36. **6.72.** (b)  $\sum_{k=0}^a x^{b(a-k)} \left[\frac{k+b-1}{k, b-1}\right]_x = \left[\frac{a+b}{a, b}\right]_x$  (look at area in Figure 2.2). **6.74.** (a)  $1728 + 864x^2 + 864x^3 + 288x^5$ . **6.75.** (d)  $10(3+x)^5$ . **6.76.**

$G_{T_4}(x) = 16x^3 + 15x^4 + 6x^5 + x^6$ . **6.77.**  $f_6(341265) = (0, 0, 2, 2, 0, 1)$  with both weights 5;  $g_6(0, 0, 1, 3, 2, 3) = 416532$  with both weights 9. **6.79.** For  $w \in S_n$ , define  $f(w) = (t_1, \dots, t_n)$  where  $t_k = |\{(i, j) \in \text{Inv}(w) : w_j = n + 1 - k\}|$ . **6.80.**  $f(35261784) = (0, 1, 1, 0, 2, 1, 3, 2)$  and  $f^{-1}(0, 1, 0, 3, 2, 4, 6, 5) = 68537142$ . **6.82.**  $f_6(341265) = (0, 0, 1, 1, 0, 5)$  with both weights 7;  $f_6^{-1}(0, 0, 1, 3, 2, 3) = 635142$  with both weights 9. **6.85.** Both sums equal  $[n]!_x$ . **6.87.** Adapt the proof of 6.29. **6.89.** (a) If  $w_n = n$ , the cyclic shift creates a new descent at position  $n - 1$  and does not affect the other descents. If  $w_1 = n$ , the cyclic shift removes 1 from the descent set. If  $w_k = n$  where  $1 < k < n$ , the cyclic shift replaces  $k$  by  $k - 1$  in the descent set. (b) Apply the cyclic shift  $n - k$  times, and then erase  $w_n = n$ . **6.90.** (b) 19,144. **6.91.**  $h_3(w) = 245331524515132$ , so inversions decrease by 6 =  $n_{>3}(w)$ . **6.92.**  $h_2^{-1}(w) = 425331542511523$ . **6.93.** e.g.,  $g(3124) = 1324$  and  $g(1432) = 4312$ . **6.94.** (a) 4526173. **6.95.** (a) 2415673. **6.96.** e.g.,  $g(100110) = 100100$ . **6.97.** (a)  $G_{W_0} = 1$ ,  $G_{W_1} = x + 1$ ,  $G_{W_n} = G_{W_{n-1}} + xG_{W_{n-2}}$  for  $n \geq 2$ , so  $G_{W_6}(x) = 1 + 6x + 10x^2 + 4x^3$ . **6.98.**  $G_{n,k} = G_{n-1,k-1} + G_{n-1,k}[k]_x$  (with appropriate initial conditions). **6.100.** 3.47 and 6.54 may be helpful here. **6.102.** Show  $\prod_{i=1}^n (1 + tx^i) = \sum_{k=0}^n t^k x^{k(k+1)/2} [n]_x$ . **6.103.** (b) Show  $(i, j) \in \text{Inv}(w)$  iff  $(w_j, w_i) \in \text{Inv}(w^{-1})$  for all  $i, j \leq n$ . (c)  $G_3(x, y) = 1 + xy + xy^2 + x^2y + x^2y^2 + x^3y^3$ . To prove  $G_n(x, y) = G_n(y, x)$ , use (a). **6.104.** (b) Let  $I(w) = w^{-1}$  as in 6.103, and consider the composition  $I \circ g \circ I \circ g^{-1} \circ I$ . **6.105.**  $G_n(x) = x^{n(n-1)/2} C_n(x^{-1})$ . **6.106.** (a)  $C_3(q, t) = q^3 + q^2t + qt^2 + t^3 + qt$ . (d) See [64] or [85]. (e) This is known to be true, using hard results in [48], but at this time (July 2010) no direct bijective proof is known. **6.107.** (a)  $k^{-1}$  sends a path  $P \in D_n$  to  $(g_0, \dots, g_{n-1})$ , where  $g_i$  is the number of area cells to the right of the path in the  $i$ th row from the bottom; check that this sequence is in  $G_n$ . (b) Map  $g \in G_n \cap \mathcal{R}(0^{v_0}1^{v_1} \dots s^{v_s})$  to a Dyck path whose bounce path has vertical moves  $v_0, v_1, \dots, v_s$ .

**7.105.**  $f + g = 1 - x - x^2 + 6x^4$ ,  $fg = x - 3x^2 + 2x^3 + 3x^4 - 3x^5 - 3x^6 + 9x^8$ ,  $\deg(f) = 4 = \deg(g) = \deg(f + g)$ ,  $\deg(fg) = 8$ ,  $\text{ord}(f) = 1 = \text{ord}(fg)$ ,  $\text{ord}(g) = 0 = \text{ord}(f + g)$ . **7.107.**  $P_g(\sqrt{5}) = 6\sqrt{5}$ ,  $P_f(x) = x^2 + 4x - 1 = f$ ,  $P_f(g) = -1 + 4x + x^2 + 4x^3 + 2x^4 + x^6$ . **7.108.** (b)  $x + x^2 + x^3/3 - x^5/30 - x^6/90$ . **7.109.** (a)  $f, g, f + g$  are nonzero,  $\deg(f) = n = \deg(g)$ , and  $f_n = -g_n$ . **7.110.** (c)  $1 - x + 3x^2/2 - 7x^3/3 + 11x^4/3$ . **7.113.** (a)  $-3x^{-2} - 3x^{-1} - 4 - 4x - 4x^2 - 4x^3 - \dots$ . **7.114.** (a)  $(\sinh x)_n = \chi(n \text{ is odd})/n!$ . **7.115.** additive identity axiom: let  $Z_n = 0_K$  for all  $n \in \mathbb{N}$ . Fix  $F \in K[[x]]$ . For  $n \in \mathbb{N}$ ,  $(F + Z)_n = F_n + Z_n = F_n + 0_K = F_n = (Z + F)_n$  (by additive identity axiom in  $K$ ), so  $F + Z = F = Z + F$ . **7.117.** (a) Let  $f = 2$  in  $\mathbb{Z}_4[x]$  (the ring  $\mathbb{Z}_4$  is defined in 2.150). (b) Let  $f = 1 + 2x$  in  $\mathbb{Z}_4[x]$ ; note  $f^{-1} = f$ . **7.118.** Use 7.61 and 7.92. **7.120.** (a) Use 2.79. (b) If  $R$  is finite, what is  $|R[x]|$ ? What is  $|^R R|$ ? **7.121.** (c) Since  $c = cx^0 \in K[[x]]$ , the definition 7.17 gives  $P_c(z) = c(0)z^0 = c1_R = c$ . **7.123.** (b) For all  $n \in \mathbb{N}$ ,  $(cF)'(n) = (n + 1)(cF)(n + 1) = c(n + 1)F(n + 1) = c(F'(n)) = (c(F'))(n)$ . So  $(cF)' = c(F')$ . **7.125.** (b) Fix  $F, G, H \in K[[x]]$  with  $G(0) = 0$ . Use 7.32 to choose polynomials  $f_n, g_n, h_n \in K[x]$  with  $g_n(0) = 0$  for all  $n$ ,  $f_n \rightarrow F$ ,  $g_n \rightarrow G$ , and  $h_n \rightarrow H$ . For each  $n \in \mathbb{N}$ , 7.57(c) gives  $(f_n h_n) \bullet g_n = (f_n \bullet g_n) \cdot (h_n \bullet g_n)$ . Now let  $n \rightarrow \infty$  and use 7.35 and 7.61. **7.127.** (c) For  $n \in \mathbb{N}$ ,  $[\log(1 + x)]'(n) = (n + 1) \log(1 + x)(n + 1) = (n + 1)(-1)^n / (n + 1) = (-1)^n$ . Now,  $\sum_{n \geq 0} (-1)^n x^n = (1 + x)^{-1}$  by 7.41. **7.129.** (a)  $3x + x^2 - 7x^3/3 + 2x^6$ ; (d)  $e^x - 1$ ; (g)  $\log(1 + x)$ . **7.130.** (a) Both sides have constant term zero; for all  $n > 0$ ,  $(\int F + G dx)(n) = (F + G)(n - 1)/n = F(n - 1)/n + G(n - 1)/n = (\int F dx)(n) + (\int G dx)(n)$ . (f) Set  $G = \int F dx$  and  $H = G'$ ; for all  $n \in \mathbb{N}$ ,  $H(n) = G'(n) = (n + 1)G(n + 1) = (n + 1)F((n + 1) - 1)/(n + 1) = F(n)$ . For  $\int F' dx$ , compute the constant term separately. **7.131.** Integration by parts rule: for all  $F, G \in K[[x]]$ ,  $\int F(G') dx = FG - F(0)G(0) - \int F'G dx$  (prove it by integrating 7.54(e) and using 7.130(a),(f)). **7.133.** (a) Coefficient of  $x^n$  is zero for  $n$  odd, and  $\sum_{k=0}^n \binom{n}{k} (-1)^k = \chi(n = 0)$  for  $n$  even. **7.136.** Use induction on  $k$ . **7.139.**  $F = 1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + x^{22} + \dots$ . (We will study this infinite product in Chapter 8.) **7.140.** For existence of the infinite products, use 7.41 and 7.33.



To verify the equalities, look at a fixed coefficient and reduce to finite products as in 7.42. **7.142.**  $\sum_{n \geq 0} x^n = (1-x)^{-1}$ . **7.144.** (b)  $\sqrt{1+x} = \sum_{n \geq 0} ((1/2) \downarrow_n / n!) x^n = 1 + x/2 - x^2/8 + x^3/16 - 5x^4/128 + \dots$ . **7.145.** (a)  $1 + x/2 + 11x^2/8 - 11x^3/16 + \dots$ . **7.146.** (a)  $1 + x + x^2/2 - x^4/8 + \dots$ . **7.147.**  $F = 3/(1+2x) + 7/(1-4x)$ , so  $F_n = 3 \cdot (-2)^n + 7 \cdot 4^n$ . **7.149.**  $F = x - 3/(1-x)^2 + (3/2)/(1-x)$ , so  $F_n = \chi(n=1) - 3n - 3/2$ . **7.151.** (b)  $F = \sum_n a_n x^n$  satisfies  $(1-3x)F = 2 + (3x/(1-x)^2)$ , so  $a_n = (17/4)3^n - (3/2)n - 9/4$ . **7.152.** (a)  $a_n = 4^n - 2^n$ ; (c)  $F = \sum_n a_n x^n$  satisfies  $F(1-6x+8x^2) = (1-2x)^{-1} - 1$ , so  $a_n = 2 \cdot 4^n - (n+2) \cdot 2^n$ . **7.154.**  $\sum_n a_n x^n = (x+5x^2+7x^3+3x^4)/(1+3x-2x^2-6x^3+x^4+3x^5) = (1-x)^{-2} - (1-x)^{-1}$ , so  $a_n = n$  for all  $n \geq 0$ . **7.157.** Treat the case  $L = 2$  separately. **7.159.** (a)  $\sec x = 1 + (1/2)x^2 + (5/24)x^4 + (61/720)x^6 + (277/8064)x^8 + \dots$ . (b) Use 7.133(a). (c) Use 7.128. (e) Work in the field  $K((x))$ . **7.160.** (a) One approach is to show both sides satisfy  $G'' + 4G = 0$ ,  $G_0 = 0$ ,  $G_1 = 2$ . (d) To start, note  $\exp(ix) = \sum_{n \geq 0} i^n x^n / n!$ , where  $i^{2k} = (-1)^k$  and  $i^{2k+1} = i(-1)^k$ . **7.162.** (b) Reduce to the case  $x = 1$  and use the idea in 7.41. (c) If  $x^n = 0_R = y^m$ , use the commutative binomial theorem to simplify  $(x+y)^{n+m-1}$ . **7.163.** Use 7.162. **7.164.** (a)  $x + x^3/6 + 3x^5/40 + 5x^7/112 + \dots$ ; (b)  $x - x^3/3 + x^5/5 - x^7/7 + \dots$ . **7.165.** For  $F = (1-rx)^{-1}$ ,  $F^{(k)} = r^k k! (1-rx)^{-k-1}$ , so Maclaurin's formula gives  $F_k = r^k$  for all  $k$ . **7.168.** (b)  $n!(F^*G^*)_n = n! \sum_{k=0}^n F_k^* G_{n-k}^* = n! \sum_{k=0}^n (F_k/k!)(G_{n-k}/(n-k)!) = \sum_{k=0}^n \binom{n}{k} F_k G_{n-k}$ . **7.169.** (a)  $\sum_n n^2 x^n = 1(1-x)^{-1} - 3(1-x)^{-2} + 2(1-x)^{-3}$ . (b) Multiply the series in (a) by  $(1-x)^{-1}$ . **7.170.** (b)  $(n(n+1)/2)^2$ . **7.173.** (a)  $S(m) = aS(m-1) + c(b^k)^m$  for  $m \geq 1$ ;  $S(0) = d$ . (b) Use partial fractions to solve  $S(1-ax) = (d-c) + c(1-b^k x)^{-1}$ . Treat the case  $a = b^k$  separately. The identity  $n^{\log_b a} = a^{\log_b n}$  can be useful here. **7.174.** Argue that  $T(1) = d$  and  $T(n) = 2T(n/2) + cn$  for  $n \geq 2$  (where  $c, d$  are constants). Now apply 7.173 with  $a = b = 2$  and  $k = 1$ . **7.175.** Use 5.90(a) and 7.173, and compare the time complexity to 5.90(b). **7.176.** Write  $G = \exp(\int P dx)$ , which is well-defined with constant term 1. The solution is  $F = (c + \int QG dx)G^{-1}$ . **7.179.** (a) Transitivity of  $\sim$ : fix  $(a, b)$ ,  $(c, d)$ ,  $(e, f)$  in  $X$  with  $(a, b) \sim (c, d)$  and  $(c, d) \sim (e, f)$ , so  $ad = bc$  and  $cf = de$ . We must show  $(a, b) \sim (e, f)$ , i.e.,  $af = be$ . Now  $af(cd) = adc f = bcde = be(cd)$ . If  $c = 0$ , then  $a = e = 0$  since  $d \neq 0 \neq f$ , so  $af = be$ . Otherwise  $cd \neq 0$ , so cancellation of  $cd$  gives  $af = be$ . (d) Injectivity of  $i$ : say  $a, b \in D$  and  $i(a) = i(b)$ . Then  $a/1 = b/1$ , so  $a1 = 1b$  and  $a = b$ .

**8.36.** (a)  $(1 - kx)^{-1}$ . **8.38.**  $x^k/(1-x)^k$ . **8.39.** (a)  $(26, 16, 10, 2, 1)$ . **8.40.** (b)  $(21, 13, 7, 1^{15})$ . **8.41.** (a)  $(9, 5, 3, 1^9)$ . **8.42.** (a)  $(15, 12, 10, 8, 6, 3, 1)$ ; (d)  $(72)$ . **8.45.** (a)  $(18, 17, 16, 15, 13, 10, 8, 7)$ . **8.46.** (b)  $\prod_{i=1}^{\infty} (1 + x^{2^i})$ . **8.47.** (a) coefficient of  $x^n$  counts integer partitions of  $n$  where all parts are divisible by 5. **8.48.** (a) The identity  $1 + 3 + 5 + \dots + (2k-1) = k^2$  can be useful. **8.50.** The result holds for the tree with one vertex. For  $t = (\bullet, t_1, t_2)$ , we can assume by induction that the result holds for  $t_1$  and  $t_2$ . If these trees have  $a_1$  and  $a_2$  leaves, then  $t$  has  $a_1 + a_2$  leaves and  $(a_1 - 1) + (a_2 - 1) + 1 = a_1 + a_2 - 1$  non-leaves, so the result holds for  $t$ . **8.53.** (a) Recursively define  $g^{-1}(0) = (\bullet, \emptyset, \emptyset)$  and (for  $k > 0$ )  $g^{-1}(k u_1 \dots u_k) = (\bullet, t_1, t_2)$ , where  $t_1 = g^{-1}((k-1)u_1 \dots u_{k-1})$  and  $t_2 = g^{-1}(u_k)$ . **8.56.** With the notation used in 8.24,  $A$  has generating function  $\prod_{i \geq 1} (1 + x^i + x^{2i} + \dots + x^{(d-1)i}) = \prod_{i \geq 1} (1 - x^{di})/(1 - x^i)$ . Argue carefully that this equals  $\prod_{i: d \text{ does not divide } i} (1 - x^i)^{-1}$ , which is the generating function for  $B$ . **8.58.** (a) Area is  $\sum_{k=1}^{2n} k - \sum_{k=1}^n k = (2n)(2n+1)/2 - n(n+1)/2 = (3n^2 + n)/2$ . **8.61.** (a) 38. **8.62.** (a) Replace each  $2^{\binom{k_i}{2}}$  by  $(1+t)^{\binom{k_i}{2}}$ . **8.63.** (a)  $(1 - \sqrt{1-4x})/2x = \sum_{n \geq 0} C_n x^n$ . **8.65.** Terms of length 4 are 3000, 2100, 2010, 1200, 1110. So the coefficient of  $x^4$  in the inverse is  $R_0^3 R_3 + 3R_0^2 R_1 R_2 + R_0 R_1^3$ , which is  $-45$  in (c). **8.66.** (c)  $x/(1-ax)$ . **8.67.** (a)  $G_S = 1 + 3xG_S$ , so  $G_S = (1-3x)^{-1} = \sum_{n \geq 0} 3^n x^n$ . (c)  $G_S$  is not defined, because  $S$  with this weight is not admissible. **8.69.** (a)  $G_S = 1 + xG_S^3$ . **8.70.** (b)  $G_S = (1 - x - \sqrt{1-2x-3x^2})/2x$ . **8.72.** Argue that  $\prod_{i \text{ odd}} (1 - x^i)^{-1} \prod_{j \text{ even}} (1 + x^j) = \prod_{k \geq 1} (1 + x^k + x^{2k} + x^{3k})$ . **8.73.** Use a bijection. **8.74.** (a)  $\prod_{i \geq 0} (1 + x^{2^i+1})$ . (b) Given a self-conjugate  $\lambda$ , consider the largest  $k$

such that  $k \leq \lambda_k$ . **8.77.** Extract the coefficient of  $x^n$  in  $n! \sum_{m \geq 0} (e^x - 1)^m / m!$ . **8.79.** (a) The left side counts pairs  $(\lambda, \mu) \in \text{Par} \times \text{DisPar}$  weighted by  $x^{|\lambda|+|\mu|}$  and signed by  $(-1)^{\ell(\mu)}$ . If  $\lambda_1 > \mu_1$  or  $\mu$  is empty, obtain  $I(\lambda, \mu)$  by dropping the first part of  $\lambda$  and adding it as the new first part of  $\mu$ . If  $\lambda_1 \leq \mu_1$  or  $\lambda$  is empty, obtain  $I(\lambda, \mu)$  by dropping the first part of  $\mu$  and adding it as the new first part of  $\lambda$ . The only fixed point is  $(0, 0)$ , giving the 1 on the right side. **8.81.** (b) Coefficient of  $x^n$  is  $\sum_{\pi} R_{N_0(\pi)} R_{N_1(\pi)} \cdots R_{N_{n-1}(\pi)}$  where we sum over Dyck paths  $\pi$  ending at  $(n-1, n-1)$ , and  $N_i(\pi)$  is the number of north steps of  $\pi$  on the line  $x = i$ . **8.83.**  $e^{-x}/(1-x)$ . **8.84.**  $\exp(t(e^x - 1 - x))$ . **8.87.** This result is due to Bressoud and Zeilberger [20]. **8.88.** (a) A bijective proof can be given using 4.126; see [109].

**9.148.** (b) Yes, every  $e \in X$  works, but  $e$  is not unique. (d) No. **9.149.** Closure: given  $x, y \in G$ , write  $x = 2k + 1$  and  $y = 2m + 1$  for some  $k, m \in \mathbb{Z}$ . Then  $x \star y = x + y + 5 = 2(k + m + 3) + 1 \in G$ . Associativity: for all  $x, y, z \in G$ ,  $(x \star y) \star z = (x + y + 5) \star z = x + y + z + 10$  and  $x \star (y \star z) = x \star (y + z + 5) = x + y + z + 10$ . Commutativity: for all  $x, y \in G$ ,  $x \star y = x + y + 5 = y + x + 5 = y \star x$ . Identity: for  $e = -5 \in G$  and  $x \in G$ ,  $x \star e = x + (-5) + 5 = x = e \star x$ . Inverses: given  $x \in G$ , let  $y = -x - 10 \in G$ ; then  $x \star y = x + (-x - 10) + 5 = -5 = e = y \star x$ , so  $y$  is the inverse of  $x$  in  $G$ . **9.151.**  $z^2 = e$  implies  $z = z^{-1}$  for all  $z \in G$ . Use this in 9.8(d). **9.154.** The definition of  $\oplus$  shows that  $(x \oplus y) \oplus z = x + y + z - cn$  for some  $c \in \{0, 1, 2\}$ , and also  $(x \oplus y) \oplus z \in \mathbb{Z}_n$ . The second condition shows that  $c$  is given by the cases in the middle of (9.1). Similar analysis works for  $x \oplus (y \oplus z)$ . **9.156.** (b)  $[x, y]C_y([x, z]) = (xyx^{-1}y^{-1})y(xzx^{-1}z^{-1})y^{-1}$ . Regroup and cancel  $y^{-1}y$  and then  $x^{-1}x$  to get  $x(yz)x^{-1}(yz)^{-1} = [x, yz]$ . **9.157.** To prove  $x^{m+n} = x^m x^n$  for  $m, n \geq 0$ , fix  $m \in \mathbb{N}$ . When  $n = 0$ , both sides are  $x^m$ . Assume  $n \geq 0$  and  $x^{m+n} = x^m x^n$  is known; then  $x^{m+(n+1)} = x^{(m+n)+1} = x^{m+n}x = (x^m x^n)x = x^m(x^n x) = x^m x^{n+1}$ . Proceed similarly when  $m < 0$  or  $n < 0$ . **9.158.** (d) Fix  $g, h \in G$  with  $\phi(g) = \phi(h)$ . Then  $g = eg = R_g(e) = R_h(e) = eh = h$ , so  $\phi$  is one-to-one. **9.159.** (a) For existence, choose  $x = a^{-1}b \in G$ . For uniqueness, use left cancellation. **9.161.** (a)  $f = (1, 3, 7, 8, 6, 4, 5)(2)$ ; (b)  $f \circ g = [5, 1, 3, 7, 2, 4, 6, 8]$ . **9.163.** (a)  $gfg^{-1} = (5, 2, 6)(1, 7)(3)(4, 8)$ ; (b)  $\text{sgn}(g) = (-1)^{8-3} = -1$ ; (c)  $h$  is not unique, but we must have  $h(4) = 6$ . **9.164.**  $\text{lcm}(\mu_1, \dots, \mu_k)$ . **9.167.** (a) Both sides are functions from  $X$  to  $X$  that send  $i_j$  to  $i_{j+1}$  for  $1 \leq j < k$ , send  $i_k$  to  $i_1$ , and fix all other  $x \in X$ . **9.168.** Given  $f = f_1 \cdots f_i \cdots f_j \cdots f_n$ ,  $f \circ (i, j) = f_1 \cdots f_j \cdots f_i \cdots f_n$  is obtained by switching the symbols in *positions*  $i$  and  $j$ ;  $(i, j) \circ f$  is obtained by switching the *symbols*  $i$  and  $j$  in  $f_1 \cdots f_n$ . **9.170.** Reduce to the case where  $f$  is a product of two basic transpositions. **9.171.** Argue that the only permutations  $w \in S_n$  giving nonzero terms in 9.37 map  $\{1, 2, \dots, k\}$  to itself and  $\{k+1, \dots, n\}$  to itself. **9.172.** (a)  $n! - 1$  additions and  $(n-1)n!$  multiplications (assuming multiplication by  $\text{sgn}(w)$  is free). (c) Use Gaussian elimination, keeping track of the effect of each elementary operation on  $\det(A)$ . **9.174.** Use the same formulas, with no signs. **9.175.** Define  $b_j = T(0_R, \dots, 1_R, \dots, 0_R) \in R$  where the  $1_R$  is in position  $j$ . Show that  $R$ -linearity forces  $T(v_1, \dots, v_n) = \sum_k b_k v_k$  for all  $v_k \in R$ . **9.177.**  $Ax = b$  forces  $x = A^{-1}b$ . The adjoint formula for  $A^{-1}$  gives  $x_i = \det(A)^{-1} \sum_{j=1}^n (-1)^{i+j} \det(A[j|i])b_j$ . The sum here is the Laplace expansion for  $\det(A_i)$  along column  $i$ . **9.178.**  $\det(AB) = -80 = (-2) \cdot 4 + 17 \cdot (-12) + (-12) \cdot (-8) + (-9) \cdot (-4)$ . **9.179.** Adapt the argument in 9.29. **9.182.** First,  $e_G = x^0 \in \langle x \rangle$ . Second, given  $y, z \in \langle x \rangle$ , write  $y = x^m$  and  $z = x^n$  for some  $m, n \in \mathbb{Z}$ . Then  $yz = x^m x^n = x^{m+n} \in \langle x \rangle$  since  $m+n \in \mathbb{Z}$ . Third, given  $y = x^m \in \langle x \rangle$ ,  $y^{-1} = x^{-m} \in \langle x \rangle$  since  $-m \in \mathbb{Z}$ . **9.184.** Imitate the proof of 9.59. **9.185.** (c) Take  $G = S_3$ ,  $S = \langle (1, 2) \rangle$ ,  $T = \langle (2, 3) \rangle$ ; note  $ST = \{e_{S_3}, (1, 2), (2, 3), (1, 2, 3)\}$  is not a subgroup since  $(1, 2, 3)^{-1} \notin ST$ . **9.186.** Consider the surjection  $f : S \times T \rightarrow ST$  defined by  $f(s, t) = st$ . For each fixed element  $z = st \in ST$ , show that  $f(u, v) = z$  iff  $u = sw$  and  $v = w^{-1}t$  for some  $w \in S \cap T$ . **9.188.** Assume  $H \leq G$  is not normal in  $G$ ; then there exist  $g \in G$ ,  $h \in H$  with  $ghg^{-1} \notin H$ . Since the unique conjugacy class of  $G$  containing  $h$  intersects both  $H$  and  $G \sim H$ ,  $H$  cannot be a union of conjugacy classes. The converse is

similar. **9.190.** The sizes of the conjugacy classes of  $S_5$  are 1, 10, 15, 20, 20, 24, and 30. A normal subgroup of  $S_5$  must be a union of conjugacy classes that includes  $\{e_{S_5}\}$  and whose size divides 120. Check that the only possibilities are  $\{e\}$ ,  $A_5$ , and  $S_5$ . **9.191.** For each  $x \in H$ , consider the sequence  $(x, x^2, x^3, \dots)$ . **9.193.** (a)  $\{[1, 2, \dots, n], [n, n-1, \dots, 2, 1]\}$ ; (b)  $S_n$ ; (c)  $S_n$ ; (d)  $\langle(1, 2, \dots, n)\rangle$ . **9.195.** (a)  $Q_k$  has  $2^k$  vertices and  $\deg(Q_k) = k$ , so  $|E(Q_k)| = k2^{k-1}$  by 3.34. (c) One can obtain  $2^k k!$  automorphisms by permuting the  $k$  positions and switching or not switching 0 and 1 in each position. Part (b) can help prove these are all of the automorphisms. **9.197.** To state the answer, consider the sizes of the equivalence classes of the graph isomorphism equivalence relation on the set  $\{C_1, \dots, C_k\}$ . **9.198.** (b) Let  $L \leq H$ . Since  $f(e_G) = e_H \in L$ ,  $e_G \in f^{-1}[L]$ . Fix  $x, y \in f^{-1}[L]$ . Then  $f(x) \in L$ ,  $f(y) \in L$ , so  $f(xy) = f(x)f(y) \in L$  and  $xy \in f^{-1}[L]$ . Also  $f(x^{-1}) = f(x)^{-1} \in L$ , so  $x^{-1} \in f^{-1}[L]$ . If  $L$  is normal in  $H$ , then  $f^{-1}[L]$  is normal in  $G$ . For if  $x \in f^{-1}[L]$  and  $g \in G$ ,  $gxg^{-1} \in f^{-1}[L]$  because  $f(gxg^{-1}) = f(g)f(x)f(g)^{-1} \in L$ . Taking  $L = \{e_H\}$  shows  $\ker(f) \trianglelefteq G$ . **9.199.** Show that the map  $z \mapsto (|z|, z/|z|)$  is a group isomorphism. **9.201.** (b) If  $H \trianglelefteq G$  and  $K \trianglelefteq G$  and  $H \cap K = \{e_G\}$ , first prove that  $hk = kh$  for all  $h \in H$  and  $k \in K$ . **9.202.** (b)  $\text{Aut}(\mathbb{Z}) = \{\text{id}_{\mathbb{Z}}, N\}$ , where  $N(k) = -k$  for all  $k \in \mathbb{Z}$ . **9.204.** (c) For  $k > 0$ , explain why the least  $m \in \mathbb{N}^+$  with  $(x^k)^m = e_G$  must satisfy  $km = \text{lcm}(k, n)$ . (d) If  $f$  is only a homomorphism, show the order of  $f(x)$  divides the order of  $x$  (when  $x$  has finite order). **9.205.** (a) Fix  $x_1, x_2, y_1, y_2 \in G$  with  $x_1H = x_2H$  and  $y_1H = y_2H$ . We know  $x_2^{-1}x_1 \in H$  and  $y_2^{-1}y_1 \in H$  and must show  $x_1y_1H = x_2y_2H$ . For this, note  $(x_2y_2)^{-1}(x_1y_1) = y_2^{-1}x_2^{-1}x_1y_1 = (y_2^{-1}y_1)(y_1^{-1}[x_2^{-1}x_1]y_1)$ . The second parenthesized expression is in  $H$  by normality, so  $(x_2y_2)^{-1}(x_1y_1) \in H$  by closure. When checking the group axioms, note  $eH = H$  is the identity of  $G/H$ , and  $xH$  has inverse  $x^{-1}H$  for  $x \in G$ . **9.207.** First show  $\bar{f}$  is well-defined: i.e., for all  $x, y \in G$ ,  $xK = yK$  implies  $\bar{f}(xK) = \bar{f}(yK)$ . **9.209.** To obtain  $f$ , apply the fundamental homomorphism theorem 9.207 to a suitable map. **9.210.** Consider  $f : C/A \rightarrow C/B$  given by  $f(xA) = xB$  for  $x \in C$ . Check  $f$  is a well-defined group homomorphism with image  $C/B$  and kernel  $B/A$ , and apply 9.207. **9.211.** 9.198 and 9.210 can be useful here. **9.213.** Closure: for  $g, x \in G$ ,  $g*x = gxg^{-1} \in G$  by the closure axiom for the group  $G$ . Identity: for  $x \in G$ ,  $e*x = exe^{-1} = xe = x$ . Associativity: for  $g, h, x \in G$ ,  $g*(h*x) = g(hxh^{-1})g^{-1} = (gh)x(gh)^{-1} = (gh)*x$ . **9.214.** Associativity: for  $k, m \in K$  and  $x \in X$ ,  $k \bullet (m \bullet x) = k \bullet (f(m)*x) = f(k)*(f(m)*x) = (f(k)f(m))*x = f(km)*x = (km) \bullet x$ . **9.216.** Use the fact that two linear maps from  $V$  to  $V$  are equal if they agree on a basis of  $V$ . **9.218.**  $H*x = \{xh^{-1} : h \in H\}$ . Since  $H$  is a group,  $h^{-1}$  ranges over  $H$  as  $h$  ranges over  $H$ . So  $H*x = \{xk : k \in H\} = xH$ . **9.220.** One approach is to note that  $\bigcap_{x \in X} \text{Stab}(x)$  is the kernel of the permutation representation associated to the given action. **9.221.** (b) Given  $g \in G$ , let  $C_g$  be the inner automorphism such that  $C_g(x) = gxg^{-1}$  for  $x \in G$ , and let  $T : G \rightarrow G$  be any automorphism. Verify that  $T \circ C_g \circ T^{-1} = C_{T(g)}$ . **9.222.** For  $(a, b) \neq (0, 0)$ , the orbit of  $(a, b)$  is the circle with equation  $x^2 + y^2 = a^2 + b^2$ , and the stabilizer of  $(a, b)$  is  $\langle 2\pi \rangle$ . **9.224.** Assume  $z \in Gx \cap Gy$ , so  $z = g*x = h*y$  for some  $g, h \in G$ . For any  $w \in Gx$ ,  $w = k*x$  for some  $k \in G$ . Check that  $w = (kg^{-1}h)*y \in Gy$ , so  $Gx \subseteq Gy$ . One proves  $Gy \subseteq Gx$  similarly. **9.227.** Fix  $x, y \in G$  with  $xH = yH$ . Then  $y^{-1}x \in H$  (by left coset equality theorem), so  $y^{-1}(x^{-1})^{-1} \in H$ , so  $Hx^{-1} = Hy^{-1}$  (by right coset equality theorem), so  $T(xH) = T(yH)$ . This shows  $T$  is well-defined. Reversing the steps shows  $T$  is one-to-one. Since  $Hy = T(y^{-1}H)$  for  $y \in G$ ,  $T$  is onto. **9.232.**  $z_{(6)} = 6$ ,  $z_{(5,1)} = 5$ ,  $z_{(2,2,2)} = 48 = z_{(2,1,1,1,1)}$ , etc. **9.233.**  $|C_{S_8}(g)| = 24$ ; conjugacy class has size 1680. **9.235.**  $Z(S_2) = S_2$ ,  $Z(S_n) = \{e_{S_n}\}$  for all  $n \neq 2$ . **9.237.** (c) Conjugacy classes are  $\{e\}$  (size 1); all 3-cycles (size 20); permutations of type  $(2, 2, 1)$  (size 15); class of  $(1, 2, 3, 4, 5)$  (size 12); class of  $(2, 1, 3, 4, 5)$  (size 12). **9.238.** Study the proof of 9.136. **9.239.** (a) 5. **9.241.**  $G$  is the disjoint union of its conjugacy classes. The term  $|Z(G)|$  counts conjugacy classes of size 1, and the sum counts the remaining conjugacy classes (by 9.130). **9.242.** Apply 9.241. **9.243.** Let  $H = \langle(1, 2, \dots, p)\rangle \leq S_p$ , and study the  $H$ -set  $S_p/H$ . **9.244.** Treat

odd  $n$  and even  $n$  separately. **9.246.** For even  $n$ , the answer is  $(k^n + k^{(n/2)})/2$ . **9.247.** (b)  $(q^9 + 6q + 2q^3 + 9q^5)/18$ . **9.248.**  $(7^4 + 11 \cdot 7^2)/12 = 245$ . **9.250.**  $(q^6 + 3q^4 + 12q^3 + 8q^2)/24$ . **9.252.**  $\binom{5}{3} = 10$ . **9.254.** (a) The symmetry group is  $A_4$ , so the answer is the coefficient of  $x_1^2 x_2^2$  in  $(p_{(1,1,1,1)} + 8p_{(3,1)} + 3p_{(2,2)})/12$ , which is 1. (b) 2. **9.256.** (a) 3; (b) 8. **9.258.** 32, 885, 748, 000 (use inclusion-exclusion).

**10.135.** Horizontal strips are  $(6)/(3)$ ,  $(5,1)/(3)$ ,  $(5,1)/(2,1)$ ,  $(4,2)/(3)$ ,  $(4,2)/(2,1)$ ,  $(4,1,1)/(2,1)$ ,  $(4,1,1)/(1,1,1)$ ,  $(3,3)/(3)$ ,  $(3,2,1)/(2,1)$ , and  $(3,1,1,1)/(1,1,1)$ . **10.137.**  $\mu/\nu$  is a horizontal strip iff  $\mu_i \geq \nu_i \geq \mu_{i+1}$  for all  $i \geq 1$ . **10.140.** (b)  $\begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 2 & 2 & \\ \hline \end{array}$  and

$$\begin{array}{|c|c|c|} \hline 1 & 1 & 2 \\ \hline 2 & 2 & \\ \hline \end{array}$$

(c) There are sixteen tableaux. **10.142.** Informally, conjugate a tableau of shape  $\mu/\nu$  to get a tableau of shape  $\mu'/\nu'$ . Formally, define  $F : \text{SYT}(\mu/\nu) \rightarrow \text{SYT}(\mu'/\nu')$  by  $F(T) = ((i, j) \mapsto T(j, i) : (i, j) \in \mu'/\nu')$  for  $T \in \text{SYT}(\mu/\nu)$ . Check  $F$  is a bijection. **10.144.** For  $N = 2$ :  $x_1 x_2^2 + x_1^2 x_2$ . For  $N = 3$ :  $x_1 x_2^2 + x_1^2 x_2 + x_1 x_3^2 + x_1^2 x_3 + x_2 x_3^2 + x_2^2 x_3 + x_1 x_2 x_3$ .

**10.145.** (a) 16; (b) the tableaux are  $\begin{array}{|c|c|c|} \hline 1 & 1 & 3 \\ \hline 2 & 2 & \\ \hline 3 & & \end{array}$  and  $\begin{array}{|c|c|c|} \hline 1 & 1 & 2 \\ \hline 2 & 3 & \\ \hline 3 & & \end{array}$ , so the coefficient is 2; (c) zero,

since the first column cannot strictly increase. **10.147.**  $N < \max_{i \geq 1} (\mu'_i - \nu'_i)$ . **10.148.** (b)  $x_1 x_2 x_3 + x_1 x_2 x_4 + x_1 x_2 x_5 + x_1 x_3 x_4 + x_1 x_3 x_5 + x_1 x_4 x_5 + x_2 x_3 x_4 + x_2 x_3 x_5 + x_2 x_4 x_5 + x_3 x_4 x_5$ . **10.149.** (a)  $x_1^3 + x_2^3 + x_3^3 + \sum_{i \neq j} x_i x_j^2$ . **10.150.** (a)  $\binom{n}{k}$ . **10.153.** One answer is  $(9, 9, 9, 8, 4, 2, 2, 2, 2, 1)/(8, 8, 8, 4, 2, 1, 1, 1, 1, 1)$ . **10.156.**  $\{m_\mu(x_1, x_2, x_3) : \mu \in \text{Par}_3(7)\} = \{m_{(7)}, m_{(6,1)}, m_{(5,2)}, m_{(4,3)}, m_{(5,1,1)}, m_{(4,2,1)}, m_{(3,3,1)}, m_{(3,2,2)}\}$ . **10.157.** e.g., for  $k = 6$  and  $N = 1, 2, \dots, 6$ , the dimensions are 1, 4, 7, 9, 10, 11 =  $p(6)$ . **10.159.** (a) 11. **10.160.**  $f_5$  maps the tableau to the tableau with rows 1, 1, 1, 1, 2, 3; 2, 3, 3, 3, 4, 4, 4; 1, 3, 3, 4, 4, 5, 6, 6; 4, 4, 5, 5, 5, 7, 9; 5, 6, 7, 7, 8. **10.162.** e.g.,  $s_{(3,1)} = m_{(3,1)} + m_{(2,2)} + 2m_{(2,1,1)} + 3m_{(1,1,1,1)}$  and  $s_{(2,1,1,1)} = m_{(2,1,1,1)} + 4m_{(1^5)}$ . **10.164.** e.g.,  $h_{(2,1)} = s_{(3,1)/(1)} = m_{(3)} + 2m_{(2,1)} + 3m_{(1,1,1)}$ . **10.165.** For the program, 10.137(a) may be useful. **10.166.** First check that  $\leq_{\text{lex}}$  is reflexive, symmetric, and transitive on  $\text{Par}(k)$ . Given  $\mu, \nu \in \text{Par}(k)$  with  $\mu \neq \nu$ , the leftmost nonzero entry in  $\mu - \nu$  is either positive or negative, so either  $\nu \leq_{\text{lex}} \mu$  or  $\mu \leq_{\text{lex}} \nu$ . **10.168.** (b)  $\mu = (2, 2, 2, 1)$ ,  $\nu = (3, 1, 1, 1, 1)$ ;  $\mu = (3, 2, 2)$ ,  $\nu = (4, 1, 1, 1)$ ;  $\mu = (3, 3, 1)$ ,  $\nu = (4, 1, 1, 1)$ ;  $\mu = (4, 3)$ ,  $\nu = (5, 1, 1)$ . **10.169.** (a)  $(5, 4, 2, 1, 1)R(6, 4, 2, 1)R(7, 3, 2, 1)$ . **10.170.** The statement fails; e.g.,  $\mu = (2, 2, 2, 1) \leq_{\text{lex}} (3, 1, 1, 1, 1) = \nu$ , but  $\nu' = (5, 1, 1) \not\leq_{\text{lex}} (4, 3) = \mu'$ . **10.172.** To prove well-ordering, recall that  $\mathbb{N}$  is well-ordered. So the first components of  $\alpha^{(j)}$  must stabilize for  $j \geq j_1$ . Then the second components must eventually stabilize, and so on. **10.173.** For the statement about  $\deg(gh)$ , it may help to show first that

$\beta \leq_{\text{lex}} \alpha$  implies  $\beta + \gamma \leq_{\text{lex}} \alpha + \gamma$  for all  $\alpha, \beta, \gamma \in \mathbb{N}^N$ . **10.174.** (a)  $\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 & 0 \\ 3 & 2 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$ .

(b)  $\mathbf{K}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 2 & -1 & -1 & 1 & 0 \\ -1 & 1 & 0 & -1 & 1 \end{bmatrix}$ , so, e.g.,  $m_{(2,2)} = s_{(1^4)} - s_{(2,1,1)} + s_{(2,2)}$ . **10.177.**

Use 9.51. **10.178.** First show (by induction) that for all  $i \in I$ , there are scalars  $b_{ij} \in K$  with  $b_{ii} \neq 0$  and  $v_i = \sum_{j \leq i} b_{ij} w_j$ . **10.179.** For  $i = 4$ ,  $T \leftarrow 4$  has rows 1, 1, 2, 3, 4, 4, 4; 2, 4, 5, 6, 6, 6; 3, 5, 7, 8; 4, 6. **10.180.** For  $i = 2$ ,  $T \leftarrow 2$  has rows 2, 2, 2, 5, 5, 7, 7; 3, 3, 3, 6, 7, 8; 4, 4, 5, 8, 8; 5, 6, 6, 9; 6, 8, 8; 7; 8. **10.182.** (b) Shift the entire first column down one row, and put  $x$  in the 1, 1-position. **10.183.** Starting at the corner box in row 3, reverse insertion produces  $T_i$  with rows 1, 1, 2, 3, 4, 6, 6; 2, 4, 5, 6, 8; 3, 5, 7; 4, 6; and  $x_i = 4$ . **10.184.** Starting at the corner box in row 5, reverse insertion produces the tableau with rows 2, 2, 4, 5, 5, 7, 7; 3, 3, 5, 6, 7, 8; 4, 5, 6, 8, 8; 6, 6, 8, 9; 7, 8; 8; and the output value

is 3. **10.187.**  $s_{(5,4,3,1,1)} + s_{(4,4,4,1,1)} + s_{(4,4,3,2,1)} + s_{(4,4,3,1,1,1)}$ . **10.188.** Final tableau has rows 1, 1, 1, 2, 2, 3, 5, 5; 2, 2, 3, 4, 4, 6; 3, 4, 5, 6, 6; 4, 5, 7, 8; 6. New boxes are (5, 1), (4, 3), (4, 4), (3, 5), (2, 6), (1, 8) in this order. **10.191.** Final tableau has rows 1, 2, 2, 3, 5, 5, 7, 7; 2, 3, 3, 5, 6, 7; 3, 4, 5, 6, 8, 8; 4, 6, 6, 7; 5, 8, 8, 8; 6, 9; 7; 8. New boxes are (1, 8), (3, 6), (5, 4), (6, 2), (7, 1), (8, 1) in this order. **10.193.** e.g.,  $\nu = (5, 5, 5, 5, 1)$  cannot be reached; insertion of 3, 2, 1 gives the shape  $\nu = (5, 5, 4, 4, 2, 1)$ . **10.195.**  $S$  has rows 2, 3, 4, 5, 7, 7, 8; 3, 5, 5, 6, 8; 4, 6, 6, 8; 6, 8, 8, 9; 7; 8; and  $z_1 z_2 z_3 z_4 = 2357$ . **10.197.** (a)  $s_{(5,4,1)} + s_{(5,3,2)} + s_{(5,3,1,1)} + s_{(4,4,2)} + s_{(4,4,1,1)} + s_{(4,3,2,1)} + s_{(4,3,1,1,1)}$ . **10.199.** (b)  $s_{(8)} + s_{(7,1)} + s_{(6,2)} + s_{(5,3)}$ . (d) First explain why  $s_{(6,3,2,2)/(3,2)} = s_{(2,2)} h_1 h_3$ . **10.200.** (a) 1; (b) 7. **10.201.** e.g.,  $h_{(2,2)} = 6m_{(1,1,1,1)} + 4m_{(2,1,1)} + 3m_{(2,2)} + 2m_{(3,1)} + m_{(4)}$ . **10.203.** (c)  $s_{(5,3)} + s_{(4,4)} + 3s_{(4,3,1)} + s_{(5,2,1)} + s_{(4,2,2)} + 2s_{(4,2,1,1)} + s_{(3,3,2)} + 2s_{(3,3,1,1)} + s_{(3,2,2,1)} + s_{(3,2,1,1,1)}$ . **10.204.** (a) 1; (b) 0. **10.205.**  $e_{(2,2,1)} = m_{(3,2)} + 2m_{(3,1,1)} + 5m_{(2,2,1)} + 12m_{(2,1,1,1)} + 30m_{(1^5)}$ . **10.207.** Count matrices  $A$  with entries in  $\{0, 1\}$  satisfying  $\sum_j A(i, j) = \alpha_i$  for all  $i$  and  $\sum_i A(i, j) = \lambda_j$  for all  $j$ . **10.208.** (a)  $2h_1(x_1, x_2)h_2(x_1, x_2) - h_1(x_1, x_2)^3 - h_3(x_1, x_2) = 0$ . **10.211.** Use 10.172 to prove termination. **10.212.** For  $m_{(2,1)}$ ,  $\alpha = (2, 1, 0, 0)$  initially, so consider  $m_{(2,1)} - e_1 e_2$ . This is  $-3e_3$ , so  $m_{(2,1)} = e_1 e_2 - 3e_3$ . **10.213.**  $J$  is  $-(x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \neq 0$ . **10.215.** The first object maps to (3, 345, 11223). **10.217.** Interchange the symbols  $h_k$  and  $e_k$  (for all  $k$ ) everywhere in the proof of 10.88. The recursion (10.7) is unchanged by this, so the proof works. **10.218.** (b)  $h_k(x_1, \dots, x_N) = h_k(x_1, \dots, x_{N-1}) + x_N h_{k-1}(x_1, \dots, x_N)$ . **10.219.** (a) Use 2.75. **10.221.** (a) 2; (b) 4; (e) compare to 10.212. **10.222.**  $h_4(2, 3, 5) = 2261$ . **10.224.** Adapt the proof of 10.88. **10.226.**  $f$  sends the first object to  $\boxed{2} \boxed{4} \boxed{4} \boxed{4^*} \boxed{4} \boxed{5} \boxed{5}$ . **10.227.**  $I$  sends the first object to (2, 14, 3333). **10.228.**  $\sum_{n \geq 0} p_{n+1}(x_1, \dots, x_N) t^n$ . **10.229.** For  $F = \prod_{i=1}^N (1 - x_i t)$ , what is  $(dF/dt)/F$ ? **10.232.**  $g(z_1) = \left( (3, 2, 1, 1, 1), (3)(7)(2, 5, 4)(6, 8)(1), \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 6 & 4 & 2 & 4 & 4 & 6 & 4 & 6 \end{pmatrix} \right)$ . **10.234.**  $I(z_1)$  has  $w = 32754681$  and  $T = 24444666$ . **10.235.** (b) Check  $A^k$  has eigenvalues  $r_1^k, \dots, r_n^k$ , hence  $\text{tr}(A^k) = p_k(r_1, \dots, r_n)$ . (c) Use 10.98 and 10.223. **10.236.** (a)  $e_3$ ; (b)  $-p_{(3,2,1,1)}$ . **10.237.** Use 7.102 and 10.86;  $h_n$  maps to  $(-1)^n e_n$ . **10.239.** e.g., applying  $\omega$  to 10.212 gives  $\text{fgt}_{(2,1)} = h_1 h_2 - 3h_3 = -2m_{(3)} - m_{(2,1)}$ . **10.241.** Let  $T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$  and  $U = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ . Then  $\text{RSK}^{-1}(T, T) = 3412$ ,  $\text{RSK}^{-1}(T, U) = 3142$ ,  $\text{RSK}^{-1}(U, T) = 2413$ , and  $\text{RSK}^{-1}(U, U) = 2143$ . **10.242.**  $P(w)$  and  $Q(w^{-1})$  have rows 1, 2, 3, 8; 4, 5, 6; 7.  $Q(w)$  and  $P(w^{-1})$  have rows 1, 3, 4, 6; 2, 5, 7; 8. **10.243.**  $w = 57218463$  and  $v = 43861725 = w^{-1}$ . **10.247.** (a)  $\text{RSK}$  is a bijection from  $S_n$  to  $\bigcup_{\lambda \vdash n} \text{SYT}(\lambda) \times \text{SYT}(\lambda)$ . Now take cardinalities of each set and use the sum and product rules. **10.250.**  $P(w)$  has rows 1, 1, 1, 1, 3; 2, 2; 3, 3; and  $Q(w)$  has rows 1, 3, 6, 7, 8; 2, 4; 5, 9.  $\text{Des}(w) = \{1, 3, 4, 8\} = \text{Des}(Q)$ . **10.252.** (a)  $t[4]_t = t + t^2 + t^3 + t^4$ . **10.253.**  $p_{(1^4)} = s_{(1^4)} + 3s_{(2,1,1)} + 2s_{(2,2)} + 3s_{(3,1)} + s_{(4)}$ . **10.255.** (a) Matrix is  $\begin{bmatrix} 0 & 1 & 0 & 1 \\ 2 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ ;  $P$  has rows 1, 1, 2, 3, 3; 2, 3; 4;  $Q$  has rows 1, 1, 2, 3, 3; 2, 2; 4. (b) Transpose the matrix in (a), and switch  $P$  and  $Q$ . **10.257.** For a solution, see [46, §4.2]. **10.260.** Show  $A$  is invertible, and then show  $B = A^{-1}$  (where  $A^{-1}$  denotes the unique two-sided inverse of  $A$ ). **10.261.** We know  $g = \sum_{\nu} a_{\nu} f_{\nu}$  for unique scalars  $a_{\nu} \in K$ . Take the scalar product of both sides with  $f_{\mu}$ . **10.264.** Define a bijection mapping a semistandard tableau to a pair consisting of a standard tableau  $U$  and an object whose weight is one of the monomials in  $Q_{n, \text{Des}(U)}$ .

**11.73.**  $\text{wt}(w) = 27$  and  $J(w) = (2, (6, 6, 4, 3, 2, 2, 2, 1, 1))$ . **11.74.** e.g.,  $U(-1, (3, 2)) = \dots 1100 \underline{10} 100 \dots$ . **11.76.** First ask how the frontier of  $\mu'$  is related to the frontier of  $\mu$ . **11.77.** Both  $w$  and  $J(w)$  have weight  $u^0 q^{17}$ . **11.80.**  $\prod_{n \geq 1} (1 - x^n) \sum_{m \in \mathbb{Z}} (-1)^m x^{(5m^2+m)/2}$  (take  $q = x^5$  and  $u = -x^{-2}$  in 11.5). **11.83.** (b)  $\begin{pmatrix} 21 \\ 6, 7, 8 \end{pmatrix} \cdot 16 \cdot 21 \cdot 14$ . **11.84.** For  $k = 3$ , the 3-core

is 0,  $\nu^0 = (2, 2)$ ,  $\nu^1 = (3, 1)$ , and  $\nu^2 = (2, 1, 1, 1)$ . **11.85.** The 2-cores are the “staircases”  $(n, n-1, \dots, 3, 2, 1, 0)$  for all  $n \geq 0$ . **11.88.** For  $k = 5$ , the 5-core is  $(2, 2)$ ,  $\nu^0 = \nu^1 = (1, 1)$ ,  $\nu^2 = (2, 2)$ ,  $\nu^3 = \nu^4 = (2)$ . **11.89.** For  $k = 4$ ,  $\nu^0 = (1)$ ,  $\nu^1 = (3)$ ,  $\nu^2 = (1, 1)$ , and  $\nu^3 = (0)$ . **11.93.**  $a_{(6,3,1)}(x_1, x_2, x_3) = +x_1^6 x_2^3 x_3^1 + x_1^3 x_2^1 x_3^6 + x_1^1 x_2^6 x_3^3 - x_1^3 x_2^6 x_3^1 - x_1^6 x_2^1 x_3^3 - x_1^1 x_2^3 x_3^6$ . **11.95.**  $K$ -linearity of the map  $T(f) = fa_{\delta(N)}$ : for  $f, g \in \Lambda_N$  and  $c \in K$ ,  $T(f+g) = (f+g)a_{\delta(N)} = fa_{\delta(N)} + ga_{\delta(N)} = T(f) + T(g)$  and  $T(cf) = (cf)a_{\delta(N)} = c(fa_{\delta(N)}) = cT(f)$ . **11.96.** (b) For  $f \in \Lambda_N$ ,  $g \in A_N$ , and  $w \in S_N$ ,  $w \bullet (fg) = (w \bullet f)(w \bullet g) = f(\text{sgn}(w)g) = \text{sgn}(w)(fg)$ , so  $fg \in A_N$ . **11.98.**  $\text{wt}(v) = x_1^3 x_2^{12} x_3^7 x_4^2 x_5^{10} x_6^{13}$ ,  $w(v) = 625314$ ,  $\text{pos}(v) = (13, 12, 10, 7, 3, 2)$ , and  $\text{sgn}(v) = +1$ . **11.100.** (c)  $a_{(3,1,1,1,1,0)+\delta(6)} - a_{(2,2,1,1,1,0)+\delta(6)}$ . For  $N \geq 7$ , add the term  $-a_{(1^7)+\delta(N)}$ . **11.101.**  $I(v, 3) = (0410030206500 \dots, 4)$ . **11.102.** (b)  $a_{(4,4,3,0,0,0)+\delta(6)} + a_{(4,3,3,1,0,0)+\delta(6)} + a_{(3,3,3,1,1,0)+\delta(6)}$ . **11.103.**  $I(v, \{3, 4, 6\}) = (0310040205600 \dots, \{3, 5, 6\})$ . **11.104.** (a)  $\sum a_{\lambda+\delta(5)}$  for  $\lambda = 721, 631, 622, 6211, 541, 532, 5311, 5221, 442, 4411, 4321$ . **11.105.** For  $M = [1, 1, 4, 5]$ ,  $I(v, M) = (v, M)$  with  $v^* = 0300104206050 \dots$ ; horizontal strip is  $(6, 5, 4, 4, 3, 1)/(5, 5, 4, 3, 1, 1)$ . **11.109.** (a)  $s_{(6,3,2)} - s_{(5,4,2)} - s_{(3,3,2,2,1)} + s_{(3,3,2,1,1,1)}$ . **11.110.** (a) 1; (b) 0. **11.112.**  $p_{(1^6)}/45 - p_{(3,1,1,1)}/9 + p_{(5,1)}/5 - p_{(3,3)}/9$ . **11.113.** e.g.,  $p_{(2,1,1)} = s_{(4)} + s_{(3,1)} - s_{(2,1,1)} - s_{(1^4)}$  and  $s_{(2,2)} = p_{(1^4)}/12 - p_{(3,1)}/3 + p_{(2,2)}/4$ . **11.114.** (a)  $I(v, T) = (5431200 \dots, T')$  where  $T'$  has rows 2, 2, 1, 1, 1; 1, 3, 4; 3, 5. **11.117.**  $-p_{(4,1)}/4 - p_{(3,1,1)}/3 + p_{(2,2,1)}/8 + p_{(2,1,1,1)}/4 + 5p_{(1,1,1,1,1)}/24$ . **11.120.** (a)  $h_{(5,3)} - h_{(6,2)}$ . **11.121.** (b)  $e_{(3,2,1)} + e_{(5,1)} - e_{(3,3)} - e_{(4,1,1)}$ . **11.122.** (a)  $h_{(3,2,2)} + h_{(5,1,1)} - 2h_{(4,2,1)}$ . **11.124.** Think of the  $x$ -coordinate as the position on an abacus and the  $y$ -coordinate as time. **11.125.** For partitions of 4, see the solution to 10.174(b). **11.126.** Generalize the proof of 11.64. **11.127.** See [33]. **11.128.** See [87]. **11.129.** (a)  $I(v^0, T) = (5423100 \dots, T')$ , where  $T'$  is the skew tableau with word  $w(T') = 22513244322111$ . **11.130.** (a) 2. **11.132.**  $s_{(5,4,1)} + s_{(5,3,2)} + s_{(5,3,1,1)} + 2s_{(4,4,2)} + s_{(4,4,1,1)} + s_{(4,3,3)} + s_{(4,3,2,1)}$ . **11.133.** (b)  $s_{(5,2)} + s_{(4,3)}$ . **11.135.** (a) 4.

**12.90.** (b) One equivalence class is {NENENE, ENENEN}. There are three other equivalence classes, each of size 6. **12.92.** Label each point  $(x, y)$  by the integer  $x - k - my$  and thereby divide the “bad” paths into  $m$  classes. Reflections do not work for  $m > 1$ , so look for other symmetries. For a solution, see [86]. **12.94.** (b) ENNNEEEENNNENEEENEEENNN. **12.95.** (a) NENENNENENNNENNEEEENEEENE. **12.96.** Check that reflection across the line  $y = x$  gives a bijection between the event  $\{\pi : X_j(\pi) = 0\}$  and the event  $\{\pi : X_j(\pi) = 1\}$ . **12.97.**  $P(X_1 + \dots + X_n = k) = \binom{n}{k}/2^n$ , which is not a uniform distribution. **12.98.** 2.30 may be helpful. **12.99.** (a)  $1 + 6x + 7x^2 + x^3$ ; (c)  $1 + nx$ . **12.102.** (a) The multiset  $[\mu_i + i : 1 \leq i \leq n]$  is  $[n+1, n+2, \dots, 2n]$ . The multiset  $[\nu_i + i : 1 \leq i \leq n]$  is  $[2n, 2n-1, \dots, n+1]$ . These multisets are equal, so 12.10 applies. (b) One approach is to find a recursion satisfied by both  $r_k(\mu)$  and  $r_k(\nu)$ . **12.105.** For a solution, see [88]. **12.106.** The first labeled path in row 2 maps to the parking function given by  $f(1) = 3$ ,  $f(2) = f(3) = 1$ . The associated tree has edges  $\{0, 2\}$ ,  $\{0, 3\}$ , and  $\{3, 1\}$ . **12.107.** (c) Labeled path is NENNNNEEEENENNNENENEE with labels 9, 2, 3, 4, 6, 7, 5, 10, 8, 1 (from bottom). Parking function is  $f(1) = 9$ ,  $f(2) = f(3) = f(4) = f(6) = 2$ ,  $f(5) = f(10) = 7$ ,  $f(7) = 6$ ,  $f(8) = 8$ ,  $f(9) = 1$ . **12.109.** (a)  $\binom{b}{c_1, \dots, c_{a+1}}$ . **12.110.** Let  $m$  be the spot where car  $n$  parks. **12.114.** Translate the proof of 12.15 into group theory. **12.115.** 2. **12.116.** The statement is false, as shown by the groups  $\mathbb{Z}_2 \times \mathbb{Z}_2$  or  $S_3$ . **12.119.** (b) Check that  $\{\pm 1, \pm i, \pm j, \pm k\}$  is a finite subgroup of  $H^*$  that is neither cyclic nor commutative. (c) For all  $(b, c, d) \in \mathbb{R}^3$  with  $b^2 + c^2 + d^2 = 1$ , use the distributive law to check that  $(bi + cj + dk)^2 = -1$  in  $H$ . **12.121.**  $(9^2 + 9^{12} - 9^4 - 9^6)/12$ . **12.122.** (a) Those of degree at most 4 are  $x$ ,  $x+1$ ,  $x^2 + x + 1$ ,  $x^3 + x + 1$ ,  $x^3 + x^2 + 1$ ,  $x^4 + x + 1$ ,  $x^4 + x^3 + 1$ ,  $x^4 + x^3 + x^2 + x + 1$ . (b) For  $1 \leq n \leq 7$ , the answers are 2, 1, 2, 3, 6, 9, 18. **12.123.** To prove that multiplicative inverses exist in  $K$ , suppose  $f \neq 0$  is in  $K$ . By irreducibility of  $h$ ,  $\gcd(f, h) = 1$ , so that  $af + bh = 1$  for some  $a, b \in F[x]$ . Check that  $a \bmod h$ , the remainder when  $a$  is divided by

$h$ , is a two-sided inverse for  $f$  in  $K$ . **12.125.** (a) For  $y \neq 0$ , use Lagrange's theorem on  $K^*$ . (c)  $x^{16} - x = x(x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1)$ . **12.126.** (a) Note  $I(n, q) \geq (q^n - (1 + q + q^2 + \cdots + q^{n-1}))/n > 0$ . (b) Use (a) and 12.123. **12.127.** (a)  $\phi(q^n - 1)/n$ ; (b)  $x$  and  $x^4 + x^3 + x^2 + x + 1$  are the two examples of smallest degree. **12.128.** (a)  $q^{n(n+1)/2}$ . **12.130.** Build such a matrix one row at a time. Each row must be chosen outside the span of the preceding rows. **12.131.**  $\begin{bmatrix} 5 \\ 3 \end{bmatrix}_7 = 140,050$ . **12.133.** RREF basis is  $(1, 4, 0, 4, 0)$ ,  $(0, 0, 1, 2, 0)$ ,  $(0, 0, 0, 0, 1)$ . **12.136.** (a) Set up a bijection between certain subspaces of  $V$  and certain subspaces of the dual space  $V^*$ . (b) Use 2.79. **12.139.** The coefficients give the probability that a randomly selected permutation is an up-down permutation. **12.141.** (b)  $n = 0$ : the empty permutation;  $n = 2$ : 12;  $n = 4$ : 1324, 1423, 2413, 2314, 3412. **12.143.** (b)  $s_0 = s_2 = 1$ ,  $s_4 = q + 2q^2 + q^3 + q^4$ . **12.144.**  $\text{wt}(t) = x_1^2 x_3^3 x_3^2 x_4^3 x_5 x_6^4$ ,  $\text{inv}(t) = 8$ ,  $\text{sgn}(t) = +1$ ,  $t$  is not transitive since  $(2, 3), (3, 4), (4, 2) \in t$ . **12.147.** Write  $A = [x_j^{N-i}]$ . Note  $\det(A) = \prod_{i < j} (x_i - x_j) \neq 0$  since the  $x_k$ 's are distinct, so  $A$  is invertible. Now write  $p = \sum_{i=0}^N c_i x^i$  where  $c_i \in F$ , and let  $v = (c_N, \dots, c_1, c_0)$ . The hypothesis  $p(x_k) = 0$  for all  $k$  means that  $vA = 0$ , hence  $v = 0$  since  $A$  is invertible. So  $p = 0$  in  $F[x]$ . **12.149.** (b) 462; (e)  $\binom{a+b-1}{b}$ . **12.150.** Classify  $T \in \text{SYT}(\lambda)$  based on the location of  $|\lambda|$  in  $T$ , and use the sum rule. **12.153.** (a)  $f^\lambda = p! / \prod_{c \in \text{deg}(\lambda)} h(c)$ . If  $\lambda$  is not a hook, each number  $h(c)$  in the denominator is less than  $p$ . Since  $p$  is prime, the factor  $p$  in the numerator cannot cancel with anything in the denominator. (b)  $(-1)^{\ell(\lambda)-1}$ . **12.157.**  $P(w)$  has shape  $(8, 5, 5, 3, 1)$ . So the longest increasing subsequence has length 8, and the longest decreasing subsequence has length 5. **12.160.** Sketch: Let  $w_1 w_2 \cdots w_{mn+1}$  have no increasing subsequence of length  $m+1$ . For  $1 \leq i \leq mn+1$ , let  $s_i$  be the length of a longest increasing subsequence of  $w$  that begins with  $w_i$ . Argue that there must exist  $n+1$  indices  $i$  whose corresponding lengths  $s_i$  are all the same, and use this to obtain a decreasing subsequence of length  $n+1$ . **12.161.** Note  $\det(A) = \det(A^t) = \det(-A) = (-1)^N \det(A) = -\det(A)$ . So  $\det(A) = 0$  (assuming the field does not have characteristic two). **12.163.** Writing  $a_{ij}$  for  $A(i, j)$ , the Pfaffian is  $a_{12}a_{34}a_{56} - a_{12}a_{35}a_{46} + a_{12}a_{36}a_{45} - a_{13}a_{24}a_{56} + a_{13}a_{25}a_{46} - a_{13}a_{26}a_{45} + a_{14}a_{23}a_{56} - a_{14}a_{25}a_{36} + a_{14}a_{26}a_{35} - a_{15}a_{23}a_{46} + a_{15}a_{24}a_{36} - a_{15}a_{26}a_{34} + a_{16}a_{23}a_{45} - a_{16}a_{24}a_{35} + a_{16}a_{25}a_{34}$ . **12.166.** (a) (15243768, 13264857); (b) (14235768, 14235768). **12.167.** (a)  $w = (1, 3, 6, 8, 2, 5)(4, 7) = 35671842$ ; term is  $-A(1, 3)A(2, 5)A(3, 6)A(4, 7)A(1, 5)A(6, 8)A(4, 7)A(2, 8)$ . **12.168.** There are 817,991 tilings of the  $6 \times 9$  board. **12.169.** (a) 144,092. **12.171.** To prove property 3, fix  $i, j \in \{1, 2, \dots, mn\}$ , and write  $i = m(i_1 - 1) + i_2$ ,  $j = m(j_1 - 1) + j_2$  where  $1 \leq i_1, j_1 \leq n$  and  $1 \leq i_2, j_2 \leq m$ . Then the  $i, j$ -entry of  $(A_1 \otimes B_1)(A_2 \otimes B_2)$  is  $\sum_{k=1}^{mn} (A_1 \otimes B_1)(i, k)(A_2 \otimes B_2)(k, j) = \sum_{k_1=1}^n \sum_{k_2=1}^m (A_1 \otimes B_1)(i, m(k_1 - 1) + k_2)(A_2 \otimes B_2)(m(k_1 - 1) + k_2, j) = \sum_{k_1=1}^n \sum_{k_2=1}^m A_1(i_1, k_1)B_1(i_2, k_2)A_2(k_1, j_1)B_2(k_2, j_2) = (\sum_{k_1=1}^n A_1(i_1, k_1)A_2(k_1, j_1)) \cdot (\sum_{k_2=1}^m B_1(i_2, k_2)B_2(k_2, j_2)) = (A_1 A_2)(i_1, j_1) \cdot (B_1 B_2)(i_2, j_2) = (A_1 A_2) \otimes (B_1 B_2)(i, j)$ . As a special case, if  $A$  and  $B$  are invertible,  $(A \otimes B)(A^{-1} \otimes B^{-1}) = (AA^{-1}) \otimes (BB^{-1}) = I_n \otimes I_m = I_{nm}$  (and similarly in the other order), so  $A \otimes B$  is invertible with inverse  $A^{-1} \otimes B^{-1}$ . **12.174.** (a) Show both sides are polynomials in  $u$  with the same zeroes and the same leading term. (b) Set  $u = 0$  in (a). **12.175.** Compare the factors indexed by  $k$  and  $n+1-k$ . For  $n$  odd, separately evaluate the product over  $j$  when  $k = (n+1)/2$ .

---

## ***Bibliography***

---

- [1] Martin Aigner, *Combinatorial Theory*, Springer-Verlag, New York (1979).
- [2] D. André, “Développement de  $\sec x$  et  $\tg x$ ,” *C. R. Acad. Sci. Paris* **88** (1879), 965–967.
- [3] D. André, “Solution directe du problème résolu par M. Bertrand,” *C. R. Acad. Sci. Paris* **105** (1887), 436–437.
- [4] D. André, “Sur les permutations alternées,” *J. Math. Pures et Appl.* **7** (1881), 167–184.
- [5] George Andrews, *The Theory of Partitions*, Encyclopedia of Mathematics and its Applications Series, Cambridge University Press, New York (1998).
- [6] George Andrews and Dominique Foata, “Congruences for the  $q$ -secant numbers,” *Europ. J. Combin.* **1** (1980), 283–297.
- [7] George Andrews and Ira Gessel, “Divisibility properties of the  $q$ -tangent numbers,” *Proc. Amer. Math. Soc.* **68** (1978), 380–384.
- [8] Michael Atiyah and Ian Macdonald, *Introduction to Commutative Algebra*, Westview Press, Boulder, CO (1994).
- [9] Edward Bender and Donald Knuth, “Enumeration of plane partitions,” *J. Combin. Theory Ser. A* **13** (1972), 40–54.
- [10] Edward Bender and S. Gill Williamson, *Foundations of Combinatorics with Applications*, Dover, Mineola, NY (2006).
- [11] Patrick Billingsley, *Probability and Measure* (third ed.), John Wiley and Sons, New York (1995).
- [12] Garrett Birkhoff, *Lattice Theory* (rev. ed.), American Mathematical Society, New York (1949).
- [13] Kenneth Bogart, *Introductory Combinatorics*, Harcourt/Academic Press, San Diego (2000).
- [14] B. Bollobás, *Modern Graph Theory*, Graduate Texts in Mathematics **184**, Springer-Verlag, New York (1998).
- [15] Miklòs Bóna, *Combinatorics of Permutations*, Chapman and Hall/CRC, Boca Raton, FL (2004).
- [16] Miklòs Bóna, *Introduction to Enumerative Combinatorics*, McGraw-Hill Higher Education, Boston (2007).
- [17] J. A. Bondy and U. S. R. Murty, *Graph Theory*, Springer, New York (2007).
- [18] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, North Holland, New York (1976).



- [19] Nicolas Bourbaki, *Elements of Mathematics: Algebra II*, Springer-Verlag, New York (2003).
- [20] David Bressoud and Doron Zeilberger, “Bijecting Euler’s partition recurrence,” *Amer. Math. Monthly* **92** (1985), 54–55.
- [21] Richard Brualdi, *Introductory Combinatorics*, North-Holland, New York (1977).
- [22] David Callan, “Bijections for the identity  $4^n = \sum_{k=0}^n \binom{2k}{k} \binom{2(n-k)}{n-k}$ ,” available online at <http://www.stat.wisc.edu/~callan/papersother>.
- [23] Peter Cameron, *Combinatorics: Topics, Techniques, Algorithms*, Cambridge University Press, Cambridge (1994).
- [24] A. Cayley, “A theorem on trees,” *Quart. J. Math.* **23** (1889), 376–378.
- [25] K. L. Chung and K. Feller, “On fluctuations in coin-tossing,” *Proc. Nat. Acad. Sci. USA* **35** (1949), 605–608.
- [26] Louis Comtet, *Advanced Combinatorics*, D. Reidel, Dordrecht (1974).
- [27] R. Diestel, *Graph Theory*, Graduate Texts in Mathematics **173**, Springer-Verlag, New York (2000).
- [28] Peter Doyle and John H. Conway, “Division by three,” preprint, [arXiv:math/0606068](https://arxiv.org/abs/math/0606068).
- [29] David Dummit and Richard Foote, *Abstract Algebra* (third ed.), Wiley, New York (2004).
- [30] Richard Durrett, *Probability: Theory and Examples* (third ed.), Duxbury Press, Pacific Grove, CA (2004).
- [31] A. Dvoretzky and T. Motzkin, “A problem of arrangements,” *Duke Math. J.* **14** (1947), 305–313.
- [32] Ö. Eğecioğlu and J. Remmel, “Bijections for Cayley trees, spanning trees, and their  $q$ -analogues,” *J. Combin. Theory Ser. A* **42** (1986), 15–30.
- [33] Ö. Eğecioğlu and J. Remmel, “A combinatorial interpretation of the inverse Kostka matrix,” *Linear Multilinear Algebra* **26** (1990), 59–84.
- [34] Susanna Epp, *Discrete Mathematics with Applications* (third edition), Thomson-Brooks/Cole, Belmont, CA (2004).
- [35] Sen-Peng Eu, Tung-Shan Fu, and Yeong-Nan Yeh, “Refined Chung-Feller theorems for lattice paths,” *J. Combin. Theory Ser. A* **112** (2005), 143–162.
- [36] Michael Fisher and H. N. V. Temperley, “Dimer problem in statistical mechanics — an exact result,” *Philos. Mag.* **6** (1961), 1061–1063.
- [37] Dominique Foata, “Further divisibility properties of the  $q$ -tangent numbers,” *Proc. Amer. Math. Soc.* **81** (1981), 143–148.
- [38] Dominique Foata, “On the Netto inversion number of a sequence,” *Proc. Amer. Math. Soc.* **19** (1968), 236–240.
- [39] Dominique Foata and John Riordan, “Mappings of acyclic and parking functions,” *Aequationes Math.* **10** (1974), 10–22.

- [40] Dominique Foata and M.-P. Schützenberger, “Major index and inversion number of permutations,” *Math. Nachr.* **83** (1978), 143–159.
- [41] Dominique Foata and M.-P. Schützenberger, “On the rook polynomials of Ferrers relations,” *Combinatorial Theory and its Applications II*, North-Holland, Amsterdam (1970), 413–436.
- [42] J. Frame, G. Robinson, and R. Thrall, “The hook graphs of the symmetric groups,” *Canadian J. Math.* **6** (1954), 316–324.
- [43] J. Françon, “Acyclic and parking functions,” *J. Combin. Theory Ser. A* **18** (1975), 27–35.
- [44] F. Franklin, “Sur le développement du produit infini  $(1-x)(1-x^2)(1-x^3)\cdots$ ,” *C. R. Acad. Sci. Paris Ser. A* **92** (1881), 448–450.
- [45] D. Franzblau and Doron Zeilberger, “A bijective proof of the hook-length formula,” *J. Algorithms* **3** (1982), 317–343.
- [46] William Fulton, *Young Tableaux*, Cambridge University Press, Cambridge (1997).
- [47] J. Förlinger and J. Hofbauer, “ $q$ -Catalan numbers,” *J. Combin. Theory Ser. A* **40** (1985), 248–264.
- [48] A. Garsia and J. Haglund, “A proof of the  $q, t$ -Catalan positivity conjecture,” *Discrete Math.* **256** (2002), 677–717.
- [49] Adriano Garsia and Stephen Milne, “Method for constructing bijections for classical partition identities,” *Proc. Natl. Acad. Sci. USA* **78** (1981), 2026–2028.
- [50] Adriano Garsia and Stephen Milne, “A Rogers-Ramanujan bijection,” *J. Comb. Theory Ser. A* **31** (1981), 289–339.
- [51] Ira Gessel, “Multipartite  $P$ -partitions and inner products of skew Schur functions,” *Combinatorics and Algebra, Contemp. Math.* **34** (1984), 289–317.
- [52] Ira Gessel, “Tournaments and Vandermonde’s determinant,” *J. Graph Theory* **3** (1979), 305–307.
- [53] Ira Gessel and Xavier G. Viennot, “Determinants, paths, and plane partitions,” preprint (1989), 36 pages. Available online at <http://people.brandeis.edu/~gessel/homepage/papers/pp.pdf>.
- [54] J. W. L. Glaisher, “A theorem in partitions,” *Messenger of Math.* N.S. **12** (1883), 158–170.
- [55] Jay Goldman, J. Joichi, David Reiner, and Dennis White, “Rook theory II. Boards of binomial type,” *SIAM J. Appl. Math.* **31** (1976), 618–633.
- [56] Jay Goldman, J. Joichi, and Dennis White, “Rook theory I. Rook equivalence of Ferrers boards,” *Proc. Amer. Math. Soc.* **52** (1975), 485–492.
- [57] B. Gordon, “Sieve-equivalence and explicit bijections,” *J. Combin. Theory Ser. A* **34** (1983), 90–93.
- [58] Henry Gould, *Combinatorial Identities; a standardized set of tables listing 500 binomial coefficient summations*, Morgantown, WV (1972).

- [59] R. Gould, *Graph Theory*, Benjamin/Cummings, San Francisco, CA (1988).
- [60] Ronald Graham, Donald Knuth, and Orem Patashnik, *Concrete Mathematics: a foundation for computer science*, Addison-Wesley, Reading (1989).
- [61] Curtis Greene, “An extension of Schensted’s theorem,” *Adv. in Math.* **14** (1974), 254–265.
- [62] Curtis Greene, Albert Nijenhuis, and Herbert Wilf, “A probabilistic proof of a formula for the number of Young tableaux of a given shape,” *Adv. in Math.* **31** (1979), 104–109.
- [63] H. Gupta, “A new look at the permutation of the first  $n$  natural numbers,” *Indian J. Pure Appl. Math.* **9** (1978), 600–631.
- [64] James Haglund, “Conjectured statistics for the  $q, t$ -Catalan numbers,” *Adv. in Math.* **175** (2003), 319–334.
- [65] James Haglund, *The  $q, t$ -Catalan Numbers and the Space of Diagonal Harmonics, with an Appendix on the Combinatorics of Macdonald Polynomials*, AMS University Lecture Series (2008).
- [66] Paul Halmos, *Naive Set Theory*, Springer-Verlag, New York (1998).
- [67] F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA (1969).
- [68] Paul Hoel, Sidney Port, and Charles Stone, *Introduction to Probability Theory*, Houghton Mifflin, Boston (1971).
- [69] Kenneth Hoffman and Ray Kunze, *Linear Algebra* (second ed.), Prentice Hall, Upper Saddle River, NJ (1971).
- [70] Thomas Hungerford, *Algebra*, Springer-Verlag, New York (1980).
- [71] Nathan Jacobson, *Basic Algebra I*, Dover paperback reprint of 1985 edition, Mineola, NY (2009).
- [72] G. James and A. Kerber, *The Representation Theory of the Symmetric Group*, Addison-Wesley, Reading, MA (1981).
- [73] J. Joichi and D. Stanton, “Bijective proofs of basic hypergeometric series identities,” *Pacific J. Math.* **127** (1987), 103–120.
- [74] Irving Kaplansky and John Riordan, “The problem of the rooks and its applications,” *Duke Math. J.* **13** (1946), 259–268.
- [75] P. W. Kasteleyn, “The statistics of dimers on a lattice I. The number of dimer arrangements on a quadratic lattice,” *Physica* **27** (1961), 1209–1225.
- [76] G. Kirchhoff, “Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird,” *Ann. Phys. Chem.* **72** (1847), 497–508.
- [77] Donald Knuth, *The Art of Computer Programming*, Volume 3 (multiple fascicles), Addison-Wesley, Reading, MA (1973).
- [78] Donald Knuth, *The Art of Computer Programming*, Volume 4 (multiple fascicles), Addison-Wesley, Reading, MA (2005).

- [79] Donald Knuth, "Permutations, matrices, and generalized Young tableaux," *Pacific J. Math.* **34** (1970), 709–727.
- [80] Wolfram Koepf, *Hypergeometric Summation*, Vieweg Verlag, Wiesbaden, Germany (1998).
- [81] A. G. Konheim and B. Weiss, "An occupancy discipline and applications," *SIAM J. Applied Math.* **14** (1966), 17–76.
- [82] Peter Lancaster, *Theory of Matrices*, Academic Press, New York (1969).
- [83] Rudolf Lidl and Harald Niederreiter, *Finite fields* (second ed.), Cambridge University Press, New York (1997).
- [84] Nicholas Loehr, "Abacus proofs of Schur function identities," *SIAM J. Discrete Math.* **24** (2010), 1356–1370.
- [85] Nicholas Loehr, "The major index specialization of the  $q, t$ -Catalan," *Ars Combinatoria* **83** (2007), 145–160.
- [86] Nicholas Loehr, "Note on André's reflection principle," *Discrete Math.* **280** (2004), 233–236.
- [87] Nicholas Loehr and Anthony Mendes, "Bijective matrix algebra," *Linear Algebra Appl.* **416** (2006), 917–944.
- [88] Nicholas Loehr and Jeffrey Remmel, "Rook-by-rook rook theory: bijective proofs of rook and hit equivalences," *Adv. in Appl. Math.* **42** (2009), 483–503.
- [89] Ian Macdonald, *Symmetric Functions and Hall Polynomials* (second edition), Oxford University Press, Oxford (1995).
- [90] P. MacMahon, *Combinatory Analysis*, AMS Chelsea Publishing, Providence (1983).
- [91] James McKay, "Another proof of Cauchy's group theorem," *Amer. Math Monthly* **66** (1959), 119.
- [92] G. A. Miller, "A new proof of Sylow's theorem," *Annals of Mathematics* **16** (1915), 169–171.
- [93] J. Susan Milton and Jesse Arnold, *Introduction to Probability and Statistics* (4th ed.), McGraw-Hill, Boston (2003).
- [94] Sri Gopal Mohanty, *Lattice Path Counting and Applications*, Academic Press, New York (1979).
- [95] J. Donald Monk, *Introduction to Set Theory*, McGraw-Hill, New York (1969).
- [96] John Moon, *Counting Labeled Trees*, Canadian Math. Congress, Montreal, Canada (1970).
- [97] John Moon, *Topics on Tournaments*, Holt, Rinehart, and Winston, New York (1968).
- [98] T. V. Narayana, *Lattice Path Combinatorics, with Statistical Applications*, University of Toronto Press, Toronto (1979).
- [99] C. Nash-Williams, "Decomposition of finite graphs into forests," *J. London Math. Soc.* **39** (1964), 12.

- [100] Albert Nijenhuis and Herbert Wilf, *Combinatorial Algorithms*, Academic Press, New York (1975).
- [101] Jean-Christophe Novelli, Igor Pak, and Alexander Stoyanovskii, “A direct bijective proof of the hook-length formula,” *Discrete Math. Theor. Comput. Sci.* **1** (1997), 53–67.
- [102] Igor Pak, “Partition bijections, a survey,” *Ramanujan Journal* **12** (2006), 5–75.
- [103] J. Peterson, “Beviser for Wilsons og Fermats Theoremer,” *Tidsskrift for Matematik* **2** (1872), 64–65.
- [104] Marko Petkovsek, Herb Wilf, and Doron Zeilberger,  $A = B$ , AK Peters Ltd., Wellesley, MA (1996).
- [105] H. Prüfer, “Neuer Beweis eines Satzes über Permutationen,” *Arch. Math. Phys.* **27** (1918), 742–744.
- [106] S. Ramanujan, “Proof of certain identities in combinatory analysis,” *Proc. Cambridge Philos. Soc.* **19** (1919), 214–216.
- [107] G. N. Raney, “Functional composition patterns and power series reversion,” *Trans. Amer. Math. Soc.* **94** (1960), 441–451.
- [108] Jeffrey Remmel, “Bijective proofs of formulae for the number of standard Young tableaux,” *Linear and Multilinear Algebra* **11** (1982), 45–100.
- [109] Jeffrey Remmel, “Bijective proofs of some classical partition identities,” *J. Combin. Theory Ser. A* **33** (1982), 273–286.
- [110] Jeffrey Remmel, “A note on a recursion for the number of derangements,” *European J. Combin.* **4** (1983), 371–374.
- [111] Jeffrey Remmel and Mark Shimozono, “A simple proof of the Littlewood-Richardson rule and applications,” *Discrete Math.* **193** (1998), 257–266.
- [112] A. Rényi, “Some remarks on the theory of trees,” *Magyar Tud. Akad. Mat. Kut. Int. Közl.* **4** (1959), 73–85.
- [113] John Riordan, *An Introduction to Combinatorial Analysis*, Wiley, New York (1958).
- [114] Fred Roberts, *Graph Theory and its Applications to Problems of Society*, SIAM, Philadelphia (1978).
- [115] Fred Roberts and Barry Tesman, *Applied Combinatorics*, Prentice Hall, Upper Saddle River, NJ (2005).
- [116] G. de B. Robinson, “On the representations of the symmetric group,” *Amer. J. Math.* **60** (1938), 745–760.
- [117] L. J. Rogers, “Second memoir on the expansion of certain infinite products,” *Proceedings London Math. Soc.* **25** (1894), 318–343.
- [118] Gian-Carlo Rota, “On the foundations of combinatorial theory I. Theory of Möbius functions,” *Zeitschrift für Wahrscheinlichkeitstheorie* **2** (1964), 340–368.
- [119] Joseph Rotman, *An Introduction to the Theory of Groups* (4th ed.), Springer-Verlag, New York (1995).

- [120] Bruce Sagan, “Congruences via abelian groups,” *J. Number Theory* **20** (1985), 210–237.
- [121] Bruce Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, Springer, New York (2001).
- [122] C. Schensted, “Longest increasing and decreasing subsequences,” *Canad. J. Math.* **13** (1961), 179–191.
- [123] M.-P. Schützenberger, “On an enumeration problem,” *J. Combin. Theory* **4** (1968), 219–221.
- [124] M.-P. Schützenberger, “Quelques remarques sur une construction de Schensted,” *Math. Scand.* **12** (1963), 117–128.
- [125] George Simmons, *Introduction to Topology and Modern Analysis*, Krieger Publishing Co., Malabar, FL (2003).
- [126] Douglas Smith, Maurice Eggen, and Richard St. Andre, *A Transition to Advanced Mathematics* (6th ed.), Brooks/Cole, Belmont, CA (2005).
- [127] Richard Stanley, *Enumerative Combinatorics* (2 volumes), Cambridge University Press, Cambridge (1997 and 1999).
- [128] Dennis Stanton and Dennis White, *Constructive Combinatorics*, Springer-Verlag, New York (1986).
- [129] James Sylvester, “A constructive theory of partitions, arranged in three acts, an interact, and an exodion,” *Amer. J. Math.* **5** (1882), 251–330.
- [130] J.-P. Tignol, *Galois’ Theory of Algebraic Equations*, World Scientific Publishing (2001).
- [131] Alan Tucker, *Applied Combinatorics*, John Wiley and Sons, New York (2002).
- [132] W. T. Tutte, “The dissection of equilateral triangles into equilateral triangles,” *Proc. Cambridge Philos. Soc.* **44** (1948), 463–482.
- [133] T. van Aardenne-Ehrenfest and N. G. de Bruijn, “Circuits and trees in oriented linear graphs,” *Simon Stevin* **28** (1951), 203–217.
- [134] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, Cambridge University Press, Cambridge (1992).
- [135] Xavier G. Viennot, “Une forme géométrique de la correspondance de Robinson-Schensted,” in *Combinatoire et Représentation du Groupe Symétrique, Lecture Notes in Math.* **579** (1977), Springer-Verlag, 29–58.
- [136] Douglas West, *Introduction to Graph Theory* (second edition), Prentice Hall, Upper Saddle River, NJ (2001).
- [137] H. Wielandt, “Ein Beweis für die Existenz der Sylowgruppen,” *Archiv der Mathematik* **10** (1959), 401–402.
- [138] Herbert Wilf, “A bijection in the theory of derangements,” *Math. Mag.* **57** (1984), 37–40.
- [139] Herbert Wilf, *generatingfunctionology*, Academic Press, New York (1990).

- [140] Herbert Wilf, "Sieve-equivalence in generalized partition theory," *J. Combin. Theory Ser. A* **34** (1983), 80–89.
- [141] Herbert Wilf, "A unified setting for selection algorithms II. Algorithmic aspects of combinatorics," *Ann. Discrete Math.* **2** (1978), 135–148.
- [142] Herbert Wilf, "A unified setting for sequencing, ranking, and selection algorithms for combinatorial objects," *Adv. in Math.* **24** (1977), 281–291.
- [143] R. Wilson, *Introduction to Graph Theory*, Longman, New York (1985).

## DISCRETE MATHEMATICS AND ITS APPLICATIONS

Series Editor KENNETH H. ROSEN

Bijjective proofs are some of the most elegant and powerful techniques in all of mathematics. Suitable for readers without prior background in algebra or combinatorics, **Bijjective Combinatorics** presents a general introduction to enumerative and algebraic combinatorics that emphasizes bijective methods.

The text systematically develops the mathematical tools, such as basic counting rules, recursions, inclusion-exclusion techniques, generating functions, bijective proofs, and linear-algebraic methods, needed to solve enumeration problems. These tools are used to analyze many combinatorial structures, including words, permutations, subsets, functions, compositions, integer partitions, graphs, trees, lattice paths, multisets, rook placements, set partitions, Eulerian tours, derangements, posets, tilings, and abaci. The book also delves into algebraic aspects of combinatorics, offering detailed treatments of formal power series, symmetric groups, group actions, symmetric polynomials, determinants, and the combinatorial calculus of tableaux.

**Features**

- Presents a readable yet rigorous exposition of enumerative combinatorics
- Emphasizes bijective methods and combinatorial proofs
- Requires minimal mathematical prerequisites
- Includes a careful treatment of ranking, unranking, and successor algorithms
- Provides detailed coverage of algebraic topics, such as formal power series, group actions, and symmetric polynomials, from a combinatorial viewpoint
- Contains numerous worked examples and applications as well as nearly 1,000 exercises, ranging in difficulty from routine verifications to unsolved problems
- Offers solutions, hints, or partial answers to many of the exercises at the end of the book

Lucid, engaging, yet fully rigorous, this text describes a host of combinatorial techniques to help solve complicated enumeration problems. It covers the basic principles of enumeration, giving due attention to the role of bijective proofs in enumeration theory.

K12206

ISBN: 978-1-4398-4884-5



9 781439 848845